THE PERFORMANCE OF REGIONALIZATION SCHEMES IN CLASSIFYING LAKES FOR BLACK BASS POPULATION CHARACTERISTICS

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Phillips Christian Perry

February 2016

THE PERFORMANCE OF REGIONALIZATION SCHEMES IN CLASSIFYING LAKES FOR BLACK BASS POPULATION CHARACTERISTICS

Phillips Christian Perry, Ph. D.

Cornell University 2016

Regionalization schemes are delineations of land areas into discrete management units which are relatively homogenous at a specific scale. Multiple schemes have been published, each of which organizes the landscape based on specific sets of factors. Some frequently cited examples of regionalization schemes include ecoregions and watersheds. Multiple studies have examined the usefulness of regionalization schemes in organizing aspects of aquatic biotic structure including the nutrient levels, water chemistry, water quality, and macro-invertebrate assemblages of lakes and the benthic invertebrate, amphibian and fish assemblages in streams. While schemes have been shown to group waterbodies by species assemblage type, less common are studies which investigate the usefulness of schemes in organizing water bodies for demographic characteristics of a single species. There are, however numerous reasons to suspect that schemes might organize waterbodies for single species population characteristics. For example, individuals within water bodies positioned in the same region of a scheme are likely to experience similar environmental conditions such as precipitation, temperature, surrounding vegetation and urban development, pollution, soil characteristics and underlying geology and they might respond similarly to those shared influences. In this dissertation, I have sought to determine how useful regionalization schemes are in explaining among-region variance in four population characteristics (relative abundance, relative condition, juvenile growth, and adult

growth) of the two black bass species (largemouth bass, *Micropterus salmoides*, and smallmouth bass, *Micropterus dolomieu*) in lakes across the State of New York. I first sought to determine which, if any, schemes were useful for which population metrics and then incorporated eleven different environmental covariates using mixed-effect multi-level models. I found that, depending on the bass metric, several schemes clustered lakes better than random chance alone. However, the lack of notable differences in bass metrics among regions provided a caution against using regionalization schemes as a primary lake management tool for black bass monitoring. I found that survey year and elevation were consistently important factors contributing to the partitioning of variance in black bass population characteristics. Finally, I identified a method for improving the accuracy of partitioning variance in bass metrics across a finite population of regions.

BIOGRAPHICAL SKETCH

(Phillips) Christian Perry was born in Casper, WY on March 8, 1978, but spent his formative years in the north-Denver suburb of Thornton, CO having moved there with his family when he was aged two. He received his B.A. in Biological Sciences with an emphasis in Field Biology and a minor in Outdoor Education from the University of Northern Colorado in 1999 and his M.A. in Outdoor Education from the same institution in 2001. From 2002 – 2006, Christian worked in the production of high-resolution satellite imagery for the company Space Imaging which later became Geo Eye and subsequently Digital Globe. While there, Christian received a graduate certificate in remote sensing from the University of Colorado, Boulder. Married to Alysia Nicole Struck in 2005, Christian and Alysia served as Peace Corps volunteers in Puebla, Mexico from 2007 – 2009. Christian then received his M.S. in Forestry and Natural Resources with an emphasis in Fisheries and Aquatic Sciences from Purdue University in 2011, the same year in which he joined the New York Cooperative Fish and Wildlife Research Unit at Cornell University as a Ph.D. student. While pursuing his doctoral degree, Christian has held several graduate research and teaching assistantships in the Department of Natural Resources, and with Alysia, has welcomed two wonderful daughters into the world, Sienna Rose Perry and Violet Mae Perry.

I dedicated my master's thesis to my grandparents, recognizing the legacy they left.

This one is for Sienna and Violet and the legacy they have yet to leave.

ACKNOWLEDGMENTS

Let's give credit where credit is due. Anyone who knows me well knows that I would amount to little more than a mass of quivering gelatin if it weren't for the love and support of my amazing wife, Alysia Nicole Struck Perry now coupled with the unquenchable energy and preposterous depths of bubbling joy and inspiration that I receive daily from our two wonderful daughters, Sienna Rose Perry and Violet Mae Perry. And I do mean wonderful! Each of them inspires me to no end. Family, I do not deserve you, nor do I take you for granted.

I thank my parents Larry and Freida Perry and in-laws Nancy Struck, Trevor Struck and Jan Hepp-Struck for their continual encouragement and unbroken closeness despite the miles between us. Ginja and Shane Huffman, Cheyenne Perry, Dakota Huffman, Nicholas and Lauren Struck, Zoey Struck, Jay and Dianne Moyers and the widespread members of the Phillips, Perry, and Struck clans have also kept me going in ways unique to each of them and I am deeply grateful to have you as my family.

I would like to recognize the many special friends we've made at East Shore Christian Fellowship during our time in Ithaca. You welcomed us from day one and we have grown to love and appreciate you immensely. You will be missed and your kindness will not be forgotten.

I thank my initial advisor, Bill Fisher, my first contact with Cornell University, who welcomed me into this black bass research project and continued to be a friendly mentor throughout my doctoral program.

No one can say enough positive things about Pat Sullivan, my primary advisor. His

clear-headed logical aptitude is matched only by his boundless sense of human decency. To me, he is the quintessential model of a gentleman and a scholar. Every time I've sat across the desk from him, I've been well aware of the fact that he must have a million more important things to do, yet each time he has treated my questions as if they were the most important items on his docket. I'm grateful.

With his years of boots-on-the ground field experience, most of which stem from his contribution and leadership at the venerable and renowned Cornell Biological Field Station at Shackelton Point, Oneida Lake, my co-advisor and committee member Randy Jackson has provided detailed and informed guidance following painstakingly close readings of my chapter drafts. His integrity and attention to detail has saved my research from more than a few wild misdirects. If any remain, it is not because he did not try to bring them to my attention.

The State of New York is well-served by the Warm-water Unit Leader of its Fisheries Bureau, Jeff Loukmas. Jeff sat in on my initial interview and has since been a continual source of helpful and friendly advice, guidance and encouragement to keep the research practical and applied. I'm glad to say he has also become a good friend along the way.

Committee member Evan Cooch's reputation as a premier quantitative ecologist precedes him and to the many that have come before me, I would add my humble acknowledgement that his reputation is well-deserved. He has the uncanny ability to quickly cut to the heart of a thorny ecological question and identify potential pathways and solutions, something he has done on several occasions related to this work.

I am especially grateful to Melanie Moss, Christie Sayre, Sarah Gould, Meghan Baumer, Shannon Hovencamp, and director of graduate studies Shorna Allred for ushering me through the administrative hurdles involved with ordering field supplies, gathering signatures, locating missing pieces of gear, meeting graduate school deadlines and almost always against the backdrop of a high level of anxiety emanating from a near-panicked recipient rushing at the last minute to get the thing done.

I experienced many unexpected moments of enjoyment from experiences shared with the instructors, fellow teaching assistants, and students of the two classes I assisted with as a teaching assistant: NTRES 2100: Introduction to Field Biology and NTRES 2010: Environmental Conservation. My hat is off to Joe Yavitt, Marc Goebel, Paul Rodewald, Ruth Sherman, Sally Nourani, Paul Simonin, Chuan Liao, Heidi Hendrichs, Martin Freehan, Maya Weltman-Fahs, James Burtis, Abbigael Foster, Francesca Cirafesi and Yue Li for your leadership, creativity, adaptability, selflessness, and sense of humor as we navigated each week of class.

My experience here was only made richer by the friendship and camaraderie of the friends and colleagues mentioned above along with fellow sojourners T.J. Ross, Alex Alexiades, Cat Sun, Chris Nadeau, Matt Adams, Jason Taylor, Kelly Robinson, Darragh Hare, Murodbek Laldjebaev, Amy Hetherington, Ben Marcy-Quay, Santi Saypanya, Taza Schaming, Erin Meyer-Gutbrod, Katie Kaplan, and Eleanore O'Neil. I wish each of you the very best as you move ever onward and upward.

This research was supported by the New York Department of Environmental Conservation, the New York Cooperative Fish and Wildlife Unit and by teaching assistantships from the Cornell University Department of Natural Resources.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

Geographic and ecological context of the study

Geographic setting

While the State of New York is centrally positioned within a northeastern USA biome that can generally be described as temperate broadleaf and mixed forests, the unique geographic setting of the state – its anchoring by two Laurentian Great Lakes to the north and west, the Hudson River to the east, and the Atlantic Ocean in the southeast – along with the topographic variability exemplified by the 1,600 meter Adirondacks in the north, the 1,200 meter Catskills in the south as well as the sea-level flat-lands near the Great Lakes and Atlantic Ocean provide for a landscape that is varied and diverse. New York State is home to several large rivers such as the St. Lawrence River delivering water from the Great lakes to the Atlantic Ocean along with the Hudson River, Mohawk River and thousands of smaller rivers draining the state's uplands. There are more than 83,500 kilometers of rivers and streams in the state. The glacial retreat of the most recent ice age has also left its mark on the environs of New York State, the most notable features being the large number (7,800) of glacial inland lakes. Besides the two Great Lakes, there are other notable large lakes such as Lake Champlain, shallow Oneida Lake, and the distinctive and deep Finger Lakes of New York's upstate region.

This varied landscape sets the stage for a diverse set of aquatic habitats resulting in a broad spectrum of niche-space available to organisms ranging from bacteria and single-celled algae to zooplankton, macroinvertebrates, fish, amphibians, reptiles, and mammals. The variety of fish species alone reflects the diversity of habitat available in New York. Fish species in New York range from saltwater to

freshwater (with multiple species migrating between the two), cold-water to warm-water and small rarely seen species to large game fish. There are over 165 species of freshwater fish alone.

Ecological status of black bass in New York

Largemouth bass *Micropterus salmoides* and smallmouth bass, *M. dolomieu* (family Centrarchidae) are the only two representatives in New York State of nine extant black bass species present in North America (Philipp and Ridgway 2002). Largemouth bass are found in shallow, vegetated, and still warm water habitats, whereas smallmouth bass prefer relatively cooler, deeper lakes and large streams (Smith 1986) with rocky substrate. Both species are native to North America with the early range of largemouth bass being more to the south than that of the smallmouth bass (Smith 1986). The range of these two species in New York was once restricted to waters connected to the Laurentian Great Lakes and the St. Lawrence River; however, both species are currently well represented in lake and rivers across the state due to their widespread introduction as sport fish (Smith 1986, Werner 2004). Both largemouth and smallmouth bass are opportunistic predators feeding on zooplankton and invertebrates during early life stages and then switching to prey fish, crayfish, and frogs as adults (Smith 1986, Werner 2004).

Largemouth and smallmouth bass exhibit the type III survivorship curve typical of many fish species where the proportion of surviving individuals after the eggs hatch declines precipitously over time. As such, they could be considered as r-selected species characterized by producing numerous young with low parental investment as opposed to K-selected species characterized by the production of few

young with high parental cost (Reznick et al. 2002). Winemiller and Rose (1992) also used smallmouth bass as a representative example of the "periodic" strategy of their fisheries-focused three-point alternative life history surface. This "periodic" strategy stands in contrast to the "opportunistic" and "equilibrium" strategies and is characterized by delayed maturity, large clutch sizes, and high survival rates (Diana 2004).

In northern states of the United States, largemouth bass mature around 3-4 years of age at an average length of 254-305mm (10-12 inches) (Green 1989, Mecozzi 1989) and smallmouth bass mature around 3-5 years at average lengths of 254-380mm (10-15 inches) (Wiegmann et al. 1992, 1997, Chu et al. 2006). Both species breed in the late spring and early summer when water temperatures are 15 to 20 degrees Celsius (Smith 1986, Green 1989). The males of both species build gravel nests at shallow depths and guard the nest during the swim-up stage (Smith 1986). Largemouth bass males tend to build their nests in shallower, weedier sites than smallmouth bass (Werner 2004). More than 1 female may deposit eggs in a given male's nest (Smith 1986).

The influence of invasive species on black bass is an important area of ongoing research. The Laurentian Great Lakes, for example, have undergone substantial food web changes as a result of the increase of invasive dreissenid (zebra and quagga) mussels beginning in the late 1980's (Kraft and Johnson 2000, Bossenbroek et al. 2001, Johnson et al. 2006) and the addition of invasive round gobies, *Neogobius melanostomus* beginning in the early 1990's (Diana 2004). Dreissenid mussels are filter feeders that effectively increase water clarity while shunting nutrients from the

water column to benthic regions (Vanderploeg et al. 2002, Hecky et al. 2004). They also change the substrate due to the accumulation of their discarded shells (Hecky et al. 2004). As visual predators, bass may temporarily benefit from the increased water clarity provided by dreissinid mussels; however, black bass may also suffer as a result of reductions in zooplankton caused by the dreissinid mussels (Diana 2004). Similarly, the invasive round goby provides certain benefits and disadvantages to black bass. Round gobies are preyed upon by black bass and the rapid increase of round gobies has been shown to increase growth rates of juvenile smallmouth bass and also has implications for bass survival, reproduction, and age at maturity (Steinhart et al. 2004). However, round gobies also consume the eggs of black bass and can thus have a limiting effect on black bass reproductive success (Steinhart et al. 2004). Moreover, round gobies may act as a reservoir species for viral hemorrhagic septicemia virus (VHSV), a pathogen responsible for mortality in black bass (Eckerlin et al. 2008, Eckerlin et al. 2011).

In addition to being influenced by invasive species, black bass may at times be considered nuisance or invasive species themselves. Due to their popularity as sportfish, smallmouth bass and largemouth bass have been widely distributed in ponds and lakes outside of their native ranges. The ranges of warmwater species like black bass are also predicted to expand northward with projected trends of global climate warming (Mandrak 1989, Stefan et al. 2001, Vander Zanden et al. 2004). Once established, bass populations can be nearly impossible to eradicate (Zipkin et al. 2008) and their introductions have led to declines in native fish abundance and diversity in north-temperate lakes (Vander Zanden et al. 1999, Jackson 2002). As the range of

black bass species expands northward into northern temperate lakes, there is concern about their interactions with native predators such as walleye, *Sander vitreus*, and lake trout, *Salvelinus namaycush*.

With the increase in fishing tournaments (Connelly and Knuth, 2013), there has been increased interest in the survival and sub-lethal effects of displacement of black bass from their previously held home ranges (Ridgway and Shuter 1996, Ridgway 2002, Wilde 2003, Siepker et al. 2007, Maynard 2013). Dispersal results have differed between smallmouth bass and largemouth bass with smallmouth bass more often returning to their original home ranges (Ridgway and Shuter 1996) and largemouth bass more often establishing new home ranges (Ridgway 2002). Concerns related to tournament-caught bass have ranged from air-exposure time during handling to damage from fishing gear to disruption of reproduction (Philipp et al. 1997, Wilde 2003, Siepker et al. 2007).

Black bass – environment relationships

Investigations of the relative influence of abiotic, environmental factors from multiple spatial scales on lentic fish assemblage patterns and species demographic characteristics provide us with insight into which environmental factors are generally considered to be most important to fish assemblages and populations (Beamesderfer and North 1995, Jackson et al. 2001, Long and Fisher 2005, Bonvechio and Bonvechio 2006). Below is an overview of several environmental variables known to be important to fish populations and assemblages in lakes. These variables can be divided into three general categories: physical, chemical and terrestrial.

Physical

Water temperature influences bass in several important ways. The rate at

which bass metabolize energy from their food is positively related to water temperature to a point, beyond which, bass are unable to metabolize fast enough to meet their high energy demands (Horning and Pearson 1973, Diana 2004). Temperature is also inversely related to oxygen content in lakes. If the temperature increases beyond some critical threshold, bass cannot obtain the oxygen they need to survive (Diana 2004). Like other fish, black bass in temperate lakes must also build up fat reserves in preparation for overwinter survival. This is particularly critical for young-of-the-year bass which must reach a critical size threshold in order to survive the winter months (Oliver et al. 1979, Garvey, et al. 1998). Shuter et al. (1980) showed that smallmouth bass required summer temperatures of at least 15 degrees Celsius to attain the size necessary for overwintering success. However, Fuhr et al. (2002) have demonstrated that abundance is more important than size in determining overwinter success for young largemouth bass. In other words, the abundance of age-1 largemouth bass after the winter season is dictated by the abundance of pre-winter age-0 largemouth bass, regardless of the post-growing period size of the age-0 largemouth bass.

The structural complexity of lakes also influences bass in important ways. Aquatic macrophytes stabilize the substrate, prevent suspension of sediments, improve water quality and clarity, and provide cover which young bass use to escape predation and adult bass make use of in capturing their prey (Moxley and Langford 1982, Savino and Stein 1982, Smart et al. 1998). Durocher et al. (1984) evaluated survey data from 30 Texas reservoirs and found a highly significant positive relationship between percent submerged vegetation and largemouth bass densities. However, if vegetative

cover is too dense, bass growth will decline because they cannot access their prey or their prey cannot remain in the vegetation. Hence, removing dense macrophytes has resulted in improvements in largemouth bass growth and size structure (Olson et al. 1998). In contrast, if vegetative cover is too sparse, bass may decimate their prey resources and then starve (Colle and Shireman 1980, Hoyer and Canfield Jr. 1996, Bonvechio and Bonvechio 2006, Middaugh 2011). Coarse woody debris such as fallen trees provides shelter and supports a prey base for bass. Near-shore substrate is also important in that both smallmouth bass and largemouth bass build their nests in shallow littoral areas. Olson et al. (2003) demonstrated that juvenile smallmouth bass were better adapted to habitats with cobble substrate and largemouth bass juveniles were more suited to habitats with more vegetation.

Chemical

As mentioned above, dissolved oxygen levels are important determinants of bass survival with depletion of oxygen often cited as the cause of massive fish kills (Moore 1942, Cooper and Washburn, 1949, Barica and Mathias, 1979). Black bass are especially sensitive to low oxygen levels during their early life stages (Siefert et al. 1974).

The acidity and alkalinity of lakes are also important to bass (Dillon et al. 1984). There are lower limits of pH in lakes beyond which largemouth and smallmouth bass do not occur (Beamish et al. 1975, Beamish 1976, Rahel and Magnuson 1983, Snuckins and Shuter 1991) and as acidity increases, bass may halt reproduction (Beamish 1976). Smallmouth bass in particular appear to be sensitive to increases in acidity in lakes (Snukins and Shuter 1991).

The abundance, condition, and growth of lentic bass populations are largely influenced by lake productivity levels. There are multiple variables which act as effective gauges of lake productivity. These variables include water clarity, light penetration depth (Secchi depth), conductivity, phosphorus and nitrogen content, and chlorophyll levels. DiCenzo et al. (1995) found that the growth and condition of spotted bass*, Micropterus punctulatus henshalli* – a congener of largemouth bass and smallmouth bass – in 10 Alabama reservoirs were associated with higher levels of chlorophyll a and conductivity and lower secchi depths. Allen et al. (1999) showed that density and growth of larval largemouth bass increased with increasing levels of chlorophyll a across 10 Alabama impoundments. Hoyer and Canfield, Jr. (1996) demonstrated that age-1 and age-2 largemouth bass growth was positively associated with total phosphorus, chlorophyll a, and total organic nitrogen concentrations across 56 Florida lakes. Bonvechio and Bonvechio (2006) showed that decreases in lake productivity were associated with declines in black crappie, *Pomoxis nigromaculatus* angler catch rates, but with increases in largemouth bass angler catch rates. Electrofishing surveys in that study also showed a significant positive relationship between memorable-sized (>508 mm TL) largemouth bass abundance and total phosphorus, but did not show significant relationships between largemouth bass of small or all sizes and productivity measures.

Terrestrial

Changes in the terrestrial environment surrounding a lake can result in indirect or direct impacts on fish populations (Schindler et al. 2000, Scheuerell and Schinidler 2004). In particular, multiple impacts to fish populations have been associated with

the conversion of forest or agricultural areas to urban land-cover. For example, increases in urbanization have been associated with increased nutrient loading (Carpenter et al. 1998), lake substrate modifications (Beauchamp et al. 1994), removal of aquatic vegetation (Ostendorp et al. 1995) and coarse woody debris (Christensen et al. 1996), the introduction of non-native species (Jackson et al. 2001), as well as increased angling pressure (Goedde and Coble 1981). These changes in the terrestrial environment result in changes in the physical and chemical makeup of lakes which then affects bass populations as shown above.

Study Objectives

In this dissertation, I have sought to shed light on how amenable New York lakes may be to classification of black bass population metrics by a variety of regionalization schemes. Below, I describe the motivation for using schemes to classify lakes, then provide support for why black bass species may be well-suited to these classifications. Finally, I describe the objectives of this dissertation.

Rationale for characterizing diverse lakes with classification schemes

There is growing interest in grouping lakes by their chemical, physical and biological characteristics to improve the efficiency of fisheries monitoring and management (Dolman 1990, Emmons et al. 1999, Wehrly et al. 2012). Both abiotic and biotic factors have been used to classify lakes and multiple statistical approaches have been employed with advantages and limitations being attributed to each (Emmons et al. 1999). In a fisheries resources context, lakes have been grouped based upon their fish assemblage patterns and then limnological conditions (water quality and lake morphology) have been associated with those groups allowing the species assemblage status of other lakes to be adequately predicted based upon known

limnological conditions (Tonn et al. 1983, Dolman 1990, Vehanen and Aspi 1996, Wherly et al. 2012). In other cases, limnological factors have been the sole variables used to group lakes (Schupp 1992, Emmons et al. 1999, Cheruvelil et al. 2008, Soranno et al. 2010, Martin et al. 2011).

Whether the primary grouping variables are biotic or abiotic, the geographic distribution of lakes within lake groups resulting from classification efforts is often not geographically contiguous, meaning lakes assigned to a single group may be positioned far apart from each other in the landscape. Geographically contiguous groupings may be conceptually appealing to managers who have seen a recent shift toward a more regional approach to biological assessment and monitoring. Geographically contiguous delineations denote relatively homogenous landscapes and the assumption of grouping lakes by these delineations would be that ecological similarities are shared among all lakes in a geographic region. Thus, the known ecological status of one lake in a region could be extrapolated to other lakes within the same region. While appealing, the utility of grouping lakes into geographically contiguous regions in order to explain variance in fish population metrics such as growth, condition, and relative abundance remains a relatively unexplored area of research. In one study, Wagner et al. (2007) evaluated the usefulness of two geographically contiguous grouping schemes – watersheds and ecoregions – in explaining variance in fish growth and found that neither scheme performed particularly well in partitioning the variability associated with mean lengths at age.

The benefit of classifications to lake fisheries managers is centered on the idea that management techniques that are effective in one lake are more likely to be

effective in other lakes of the same class. Often, there are more lakes than can be adequately sampled in a given management district and so classifying lakes can aide in managing unsampled waters based on sampling efforts in lakes that are known to be similar. Thus, lake classifications can facilitate increased efficiency in lake management. Lake classifications can also be useful in establishing baseline water quality, community assemblage, or single-species population metric conditions and can simplify long-term monitoring programs.

The case for classifying New York lakes for black bass demographic characteristics

The ability to explain variability in fish demographic characteristics by classifying lakes into regions would present fisheries resource managers with a powerful tool to efficiently monitor and manage lakes for fish conservation. In selecting fish species for investigations into how well variability in fish demographic characteristics is explained by regionalization schemes, it is important that the fish species satisfy a basic set of criteria. The species should be represented by sufficient numbers in lakes that are widely distributed across the region of interest. Collection of individuals of the species should be carried out in a standardized manner such that the estimates of demographic characteristics are comparable with minimal bias associated with them. The species should have a substantial role in influencing the structure of the ecological community to which it is a part. And management of the species demographic characteristics should be a central component of existing management objectives which are reflective of the fisheries value of the species.

All of these general criteria are satisfied by the two species of black bass (largemouth bass, and smallmouth bass) present in lakes across New York. These

species are well represented by standardized collection surveys in a statewide fisheries database maintained by the New York State Department of Environmental Conservation (NYSDEC), in extensive fisheries databases for New York's portion of Lake Erie and the eastern basin of Lake Ontario, and in a long-term fisheries database storing data on Oneida Lake. According to the 2007 angler survey (Connelly and Brown 2009), black bass are the most popular sport fish in New York State. Black bass are also a keystone predator in lake food webs (Mittelbach et al. 1995). Moreover, the two black bass species present in New York lakes represent coolwater and warmwater preferences. Thus, these species provide for a contrast in the effectiveness of using regionalization schemes to explain regional variability in demographic characteristics.

Chapter Descriptions

Chapter One

In chapter one, I sought to identify existing regionalization schemes which might efficiently cluster lakes for four black bass population metrics: relative abundance, relative weight, growth of juveniles and growth of adults. The seven schemes I chose to evaluate were selected based on both their previous use in the literature for aquatic studies and for the recent trend in natural resource management toward the use of ecoregions and catchments to manage resources. Regionalization schemes have not often been evaluated for use in classifying waterbodies for fish population demographics, but if a particular scheme performed well in clustering lakes for fish populations, it could represent a considerable savings to management agencies in terms of both money and time and would provide a template from which future monitoring strategies could be developed.

Noting that previous similar studies have included the regions of schemes as random variables in the same kinds of multi-level mixed-effect models that I used to evaluate the effectiveness of schemes, I also evaluated regions as random variables. However, acknowledging the evidence that it is more appropriate statistically to use the regions of schemes as fixed effects when their extent is limited to geographic boundaries (in my case, the boundaries of the State of New York), I also ran a separate set of multi-level, mixed-effect models where region was included as a fixed effect. Doing so highlighted a subsequent problem – the lack of applicability of the models to regions that were left unsampled – which I attempted to solve in two contrasting ways. The first way, deemed the ad-hoc method, was to further penalize the AIC scores of each model by adding the product of two times the sum of the number of regions unrepresented by data. The second method, deemed the sophisticated method, involved more complex mathematics and thus became the primary subject of chapter three.

Chapter Two

Having identified regionalization schemes which performed better than chance alone in clustering lakes for black bass population metrics and which, for some metrics explained relatively large proportions of the total variance associated with the response metric, in chapter two, I further evaluated the influence of eleven environmental covariates commonly found to be important to bass demographic characteristics. Doing so provided insight as to why some schemes may have performed better than others in organizing lakes for bass population metrics and also provided a strong link to existing literature and traditional methodology used to

evaluate bass populations in lakes. These environmental covariates included lake depth, surface area, shoreline development, elevation, latitude and longitude, and proportions of land cover types in the lake's catchment. Also included were three variables related to time and temperature: year, mean annual summer temperature, and degree days.

Chapter Three

In chapter three, I further developed the sophisticated method alluded to earlier as a solution for solving the problem of having regions from a finite population represented as random variables (which carries the implicit assumption of selection from an infinite population) in a mixed-effect multi-level model. Doing so involved the use of a finite population correction, fpc, properly inserted into the likelihood equation of the multi-level mixed-effect model. Where the focus of the ad-hoc solution mentioned in chapter one was on the penalty side of the AIC equation, the focus of the more sophisticated mathematical solution of chapter three was on the likelihood, goodness-of-fit side of the AIC equation.The solution developed in chapter three acknowledges a growing body of statistical literature involving the application of AIC model comparison techniques in general (Hurvich and Tsai 1989, Vaida and Blanchard), and specifically to mixed-effect models (Vaida and Blanchard 2005), as well as alternatives such as leave-one-cluster-out cross-validation (Fang 2011).

Description of key findings

Much of this study was predicated on the idea that regionalization schemes might provide a satisfactory method for clustering lakes for bass demographic metrics. The logic driving this idea was multifaceted. For example, bass populations in lakes situated within relatively homogenous regions in terms of landscape features should

be experiencing many similarities in such things as air temperature and precipitation, surrounding vegetation, urban development levels, pollution sources, soil characteristics and geological features. In other words, we felt regional geographical features might play a role in influencing the demographic characteristics of lacustrine bass populations. We hypothesized that lakes would cluster well by region for black bass population metrics and that some regionalization schemes would perform better than others in clustering lakes for bass population attributes. Moreover, we identified pertinent environmental variables that, when included in mixed-effect multi-level models involving regionalization schemes, could provide insight as to why certain schemes and regions clustered bass populations better than others.

Despite the fact that our results show that bass population metrics were not convincingly grouped by regionalization schemes, our modeling approaches did reveal several important relationships which provide insight as to which geographical features are most influential on bass growth, condition and abundance. For example, we found that elevation was consistently negatively related to abundance, growth and condition metrics across both species of bass suggesting that lakes that are higher in elevation tend to have reduced growth, reduced abundance, and lower body condition than those in lower elevations. The highest lake in our dataset was approximately 200 m above sea level and while this elevation is not extreme, it was sufficient to produce a notable (and in several cases statistically significant) negative influence on bass demographics. We suggest that this relationship is the result primarily of cooler temperatures in the higher elevation lakes. The numerous statistically significant positive relationships between the growth metrics of both species of bass and the mean

summer temperature averaged across the lifespan of the bass provide further evidence in support of this notion of temperature being a primary driver of bass demographic characteristics. That we also identified negative relationships between largemouth bass growth metrics and latitude of lake also suggests the important role that temperature plays in bass growth with bass in lakes that are further north (i.e., cooler higher latitudes) experiencing slower growth than their southern counterparts.

While several predictors hinted at the strong positive relationship between bass demographic characteristics and temperature, that was not the only signal we identified as to which environmental features bass populations respond most to. Our results also show that nutrient loading is a significant contributor to bass population metrics. Although we did not explicitly incorporate chemical nutrient measures as environmental predictors in our models, our inclusion of the proportion of cultivated land-use in the lake catchment acted as a surrogate variable to account for nutrient impacts. That bass condition and growth metrics were positively (and statistically significantly) related to the proportion of cultivated land-use in the lakes' catchments suggests that nutrient loading is a substantial contributor to bass population attributes.

Taken together, the importance of temperature and nutrient loading on bass demographic characteristics suggest that there may be ways of clustering lakes according to broad landscape patterns. That the regionalization schemes that we chose to examine did not provide clear and convincing clustering of lakes may be related to the fact that the regions of these schemes did not provide adequate separation across the landscape of the state of New York in terms of water body temperature and nutrient loading. In other words, any given region within our schemes might have

high within-region variation in temperature and nutrient loading and/or the between-region temperature and nutrient loadings might not be distinct enough to produce a measurable signal.

The two black bass species represented in the State of New York and included as subject species in our research are relatively tolerant to wide ranges of environmental conditions, a characteristic that has allowed them to be successfully introduced in water bodies across the United States. This physiological aspect of black bass may have worked to our disadvantage in trying to cluster them via regionalization schemes. Even though the bass populations did exhibit apparent preferences in temperature and nutrient loadings, the generally resilient nature of black bass allow their populations to prosper under relatively wide ranges of multiple environmental conditions.

References

Allen, M.S., J.C. Greene, F.J. Snow, M.J. Maceina, and D.R. DeVries. 1999. Recruitment of largemouth bass in Alabama reservoirs: relations to trophic state and larval shad occurrence. North American Journal of Fisheries Management 19:67-77.

Barica, J. and J.A. Mathias. 1979. Oxygen depletion and winterkill risk in small prairie lakes under extended ice cover. Journal of the fisheries Board of Canada 36:980-986.

Beamesderfer, R.C.P., and J.A. North. 1995. Growth, natural mortality, and predicted response to fishing for largemouth bass and smallmouth bass populations in North America. North American Journal of Fisheries Management 15:688-

704.

Beamish, R.J., W.L. Lockhart, J.C. Van Loon, H.H. Harvey. 1975. Long-term acidification of a lake and resulting effects on fishes. Ambio 4:98-102.

Beamish, R.J. 1976. Acidification of lakes in Canada by acid precipitation and the resulting effects on fishes. Water, Air, and Soil Pollution 6:501-514.

Beauchamp, D.A., E.R. Byron, and W.A. Wurtsbaugh. 1994. Summer habitat use by littoral-fishes in Lake Tahoe and the effects of shoreline structures. North American Journal of Fisheries Management 14:385-394.

Bonvechio, K.I. and T.F. Bonvechio. 2006. Relationship between habitat and sport fish populations over a 20-year period at west lake Tohopekaliga, Florida. North American Journal of Fisheries Management 26:124-133.

Bossenbroek, J.M., C.E. Kraft, and J.C. Nekola. 2001. Prediction of long-distance dispersal using gravity models: Zebra mussel invasion of inland lakes. Ecological Applications 11:1778-1788.

Carpenter S.R., N.F. Caraco, D.L. Correll, R.W. Howarth, A.N. Sharpley, and V.H. Smith. 1998. Nonpoint pollution of surface waters with phosphorus and nitrogen. Ecological Applications 8:559-568.

Cheruvelil, K.S., P.A. Soranno, M.T. Bremigan, T. Wagner, and S.L. Martin. 2008. Grouping lakes for water quality assessment and monitoring: the roles of regionalization and spatial scale. Environmental Management 41:425-440.

Christensen D.L., B.J. Herwig, D.E. Schindler, S.R. Carpenter. 1996. Impacts of lakeshore residential development on coarse woody debris in north temperate lakes. Ecological Applications 6:1143-1149.

Chu C., N.C. Collins, N.P. Lester, and B.J. Shuter, 2006. Population dynamics of smallmouth bass in response to habitat supply. Ecological Modelling 349-362.

Colle, D.E. and J.V. Shireman. 1980. Coefficients of condition for largemouth bass, bluegill and redear sunfish in hydrilla infested lakes. Transactions of the American Fisheries Society 109:521-531.

Connelly, N.A. and T.L. Brown. 2009. New York statewide angler survey 2007. New York State Department of Environmental Conservation, Bureau of Fisheries, Albany, NY.

Connelly, N.A., and B.A. Knuth. 2013. Black bass anglers' experiences in New York and their views on tournaments and fisheries management. New York State Department of Environmental Conservation, Bureau of Fisheries, Albany, NY.

Cooper, G.P. and G.N. Washburn. 1949. Relation of dissolved oxygen to winter mortality of fish in Michigan lakes. Transactions of the American Fisheries Society 76: 23-33.

Diana, J.S. 2004. Biology and Ecology of Fishes, 2nd edition. Cooper Publishing Group, Traverse City, Michigan.

Dicenzo, V.J., M.J. Maceina, and W.C. Reeves. 1995. Factors related to growth and condition of the Alabama subspecies of spotted bass in reservoirs. North American Journal of Fisheries Management 15:794-798.

Dillon, P.J., N.D. Yan, H.H. Harvey, and D.W. Schindler. 1984. Acidic deposition: Effects on aquatic ecosystems. C R C Critical Reviews in Environmental Control 13:167-194.

Dolman, W.B., 1990. Classification of Texas reservoirs in relation to limnology and

fish community associations. Transactions of the American Fisheries Society 119:511-520.

Durocher, P.P., W.C. Provine, and J.E. Kraai. 1984. Relationship between abundance of largemouth bass and submerged vegetation in Texas reservoirs. North American Journal of Fisheries Management 4:84-88.

Eckerlin, G.E. 2008. Viral Haemorrhagic Septicaemia Virus Type IVb among smallmouth bass Micropterus dolomieu (Lacepede) in the St. Lawrence River: An iinvasive species mediated pathogen? Masters thesis. State University of New York; College of Environmental Science and Forestry, Syracuse, New York, USA.

Eckerlin, G.E., J.M. Farrell, R.N. Casey, K.M. Hope, G.H. Groocock, P.R. Bowser, and J. Casey. 2011. Temporal variation in prevalence of viral hemorrhagic septicemia virus type IVb among upper St. Lawrence River smallmouth bass. Transactions of the American Fisheries Society 140:529-536.

Emmons, E.E., M.J. Jennings, and C. Edwards. 1999. An alternative classification method for northern Wisconsin lakes. Canadian Journal of Fisheries and Aquatic Sciences 56:661-669.

Fang, Y. 2011. Asymptotic equivalence between cross-validations and Akaike information criteria in mixed-effects models. Journal of data science 9:15-21.

Fuhr, M.A., D.H. Wahl, and D.P. Philipp. 2002. Fall abundance of age-0 largemouth bass is more important than size in determining age-1 year class strength in Illinois. Pages 91-99 in D.P.Philipp and M.S. Rigeway, editors. Black bass: ecology, conservation, and management. American Fisheries Society,

Symposium 31, Bethesda, Maryland.

Garvey, J.E., R.A. Wright, and R.A. Stein. 1998. Overwinter growth and survival of age-0 largemouth bass (Micropterus salmoides): revisiting the role of body size. Canadian Journal of Fisheries and Aquatic Science 55:2414-2424.

Goedde, L.E., and D.W. Coble. 1981. Effects of angling on a previously fished and an unfished warmater fish community in two Wisconsin lakes. Transactions of the American Fisheries Society 110:594-603.

Green, D.M. 1989. Centrarchid Sampling Manual. New York State Department of Environmental Conservation. Albany, NY.

Hecky, R.E., R.E.H. Smith, D.R. Barton, S.J. Guildford, W.D. Taylor, M.N. Charlton, and T. Howell. 2004. The nearshore phosphorus shunt: a consequence of ecosystem engineering by dreissenids in the Laurentian Great Lakes. Canadian Journal of Fisheries and Aquatic Science 61:1285-1293.

Horning, W.B., II, and R.E.Pearson. 1973. Growth temperature requirements and lower lethal temperatures for juvenile smallmouth bass (Micropterus dolomieui). Journal of the Fisheries Research Board of Canada 30:1227-1230.

Hoyer, M.V. and D.E. Canfield, Jr. 1996. Largemouth bass abundance and aquatic vegetation in Florida lakes: an empirical analysis. Journal of Aquatic Plant Management 34:23-32.

Hurvich, C. M., and C.L. Tsai. 1989. Regression and time series model selection in small samples. Biometrika 76:297-307.

Jackson, D.A., P.R. Peres-Neto, and J.D. Olden. 2001. What controls who is where in freshwater fish communities – the roles of biotic, abiotic, and spatial factors.

Canadian Journal of Fish and Aquatic Sciences 58:157-170.

Jackson, D.A. 2002. Ecological effects of Micropterus introductions: the dark side of black bass. American Fisheries Society Symposium 31:221-232.

Johnson, L.E, J.M. Bossenbroek and C.E. Kraft. 2006. Patterns and pathways of the post-establishment spread of non-indigenous aquatic species across multiple spatial and temporal extents: the slowing invasion of inland lakes by the zebra mussel. Biological Invasions 8:475-489.

Kraft, C.E. and L.E. Johnson. 2000. Regional differences in rates and patterns of North American inland lake invasions by zebra mussels (Dreissena polymorpha). Canadian Journal of Fisheries and Aquatic Sciences 57:993-1001.

Long, J.M. and W.L. Fisher. 2005. Distribution and abundance of black bass in Skiatook Lake, Oklahoma, after introduction of smallmouth bass and a liberalized harvest regulation on spotted bass. North American Journal of Fisheries Management 25:49-56.

Mandrak, N.E. 1989. Potential invasion of the Great Lakes by fish species associated with climate warming. Journal of Great Lakes Research 15:306-316.

Martin, S.L., P.A. Soranno, M.T. Bremigan, and K.S. Cheruvelil. 2011. Comparing hydrogeomorphic approaches to lake classification. Environmental management 48:957-974.

Maynard, G. 2013. Lake Champlain smallmouth (Micropterus dolomieu) and largemouth (M. salmoides) bass stress levels and dispersal following catch and release tournaments in Plattsburgh, NY. Masters Thesis. SUNY Plattsburgh,

Plattsburgh, NY.

Mecozzi, M., 1989. Largemouth Bass Fact Sheet. Wisconsin Department of Natural Resources Bureau of Fisheries Management PUBL-FM-709 89.

Middaugh, C. 2011. Habitat influences on juvenile largemouth bass (Micropterus salmoides) and bluegill (Lepomis macrochirus) distrubtions, diets and movement strategies in lentic systems. Masters Thesis. Purdue University.

Miller, P.J. 1986. Gobiidae. Pages 1019-1085 in P. J. P. Whitehead, M.L. Bauchot., J.C. Hureau, J. Nielsen, E. Tortonese, editors. Fishes of the north-eastern Atlantic and the Mediterranean, volume III. United Nations Educational, Scientific and Cultural Organization, Paris, France.

Moore, W.G. 1942. Field studies on the oxygen requirements of certain fresh-water fishes. Ecology 23:319-329.

Moxley, D.J. and F.H. Langford. 1982. Beneficial effects of hydrilla on two eutrophic lakes in central Flordia. Proceedings of the Annual Conference of the Southeastern Association of Fish and Wildlife Agencies 36:280-286.

Oliver, J.D., G.F. Holeton, and K.E. Chua. 1979. Overwinter mortality of fingerling smallmouth bass in relation to size, relative energy stores, and environmental temperature. Transactions of the American Fisheries Society 108:130-136.

Olson, M.H., S.R. Carpenter, P. Cunningham, S. Gafny, B.R.Herwig, N.P. Nibbelink, I. Pellett, C. Storlie, A.S. Trebitz, and K.A. Wilson. 1998. Managing macrophytes to improve fish growth: a multi-lake experiment. Fisheries 23:6-12.

Olson, M.H., B.P. Young, K.D. Blinkoff. 2003. Mechanisms underlying habitat use

of juvenile largemouth bass and smallmouth bass. Transactions of the American Fisheries Society 132:398-405.

Ostendorp, W., C. Iseli, M. Krauss, P. Krumscheid-Plankert, J. Moret, M. Rollier, and F. Schanz. 1995. Lake shore deterioration, reed management and bank restoration in some central European lakes. Ecological Engineering 5:51-75.

Philipp, D.P., C.A. Toline, M.F. Kubacki, D.B.F. Philipp and F.J.S. Phelan. 1997. The impact of catch-and-release angling on the reproductive success of smallmouth bass and largemouth bass. North American Journal of Fisheries Management 17:557-567.

Philipp, D.P., and M.S. Ridgway, editors. 2002. Black bass: ecology, conservation, and management. American Fisheries Society, Symposium 31, Bethesda, Maryland.

Rahel, F.J. and J.J. Magnuson. 1983. Low pH and the absence of fish species in naturally acidic Wisconsin lakes: inferences for cultural acidification. Canadian Journal of fisheries and Aquatic Sciences 40: 3-9.

Reznick D., M.J. Bryant, and F. Bashey. 2002. R- and K-selection revisited: the role of population regulation in life-history evolution. Ecology 83: 1509-1520.

Ridgway M.S. and B.J. Shuter. 1996. Effects of displacement on the seasonal movements and home range characteristics of smallmouth bass in Lake Opeongo. North American Journal of Fisheries Management 16:371-377.

Ridgway, M.S. 2002. Movements, home range, and survival estimation of largemouth bass following displacement. American Fisheries Society Symposium 31:525-533.

Savino, J. and R.A. Stein. 1982. Predator-prey interaction between largemouth bass and bluegills as influenced by simulated, submersed vegetation. Transactions of the American Fisheries Society 111:255-266.

Scheuerell, M.D. and Schindler, D.E. 2004. Changes in the spatial distribution of fishes in lakes along a residential development gradient. Ecosystems 7:98-106.

Schindler, D.E., S.I. Geib, and M.R. Williams. 2000. Patterns of fish growth along a residential development gradient in north temperate lakes. Ecosystems 3:229-237.

Schupp, D.H. 1992. An ecological classification of Minnesota Lakes with associated fish communities. Minnesota Department of Natural Resources Investigational Report 417:1-27.

Shuter, B.J., J.A. Maclean, F.E.J. Fry and H.A. Regier. 1980. Stochastic simulation of temperature effects on first-year survival of smallmouth bass. Transactions of the American Fisheries Society 109:1-34.

Siefert, R.E., A.R. Carlson, and L.J. Herman. 1974. Effects of reduced oxygen concentrations on the early life stages of mountain whitefish, smallmouth bass, and white bass. The progressive fish-culturist 36:186-190.

Siepker, M.J., K.G. Ostrand, S.J. Cooke, D.P. Philipp, and D.H. Wahl. 2007. A review of the effects of catch-and-release angling on black bass, Micropterus spp.: implications for conservation and management of populations. Fisheries management and ecology 14:91-101.

Smart, R.M., G.O.Dick, and R.D. Doyle. 1998. Techniques for establishing native aquatic plants. Journal of Aquatic Plant Management 36:44-49.

Smith, C.L. 1986. The inland fishes of New York State. New York State Department of Environmental Conservation, Albany, NY.

Soranno, P.A., K.S. Cheruvelil, K.E. Webster, M.T. Bremigan, T. Wagner, and C.A. Stow. 2010. Using landscape limnology to classify freshwater ecosystems for multi-ecosystem management and conservation. Bioscience 60:440-454.

Stefan, H.G., X. Fang, and J.G. Eaton. 2001. Simulated fish habitat changes in North American lakes in response to projected climate warming. Transactions of the American Fisheries Society 130:459-477.

Steinhart, G.B. 2004. Exploring factors affecting smallmouth bass nest success and reproductive behavior. Ph.D. dissertation. The Ohio State University, Columbus, Ohio, USA.

Tonn, W.M., J.J. Magnuson and A.M. Forbes. 1983. Community analysis in fishery management: an application with northern Wisconsin lakes. Transactions of the American Fisheries Society 112:368-377.

Vaida, F., and S. Blanchard. 2005. Conditional Akaike information for mixed-effects models. Biometrika 92:351-370.

Vanderploeg, H.A., T.F. Nalepa, D.J. Jude, E.L. Mills, K.T. Holeck, J.R. Liebig, I.A. Grigorovich, and H. Ojaveer. 2002. Dispersal and emerging ecological impacts of Ponto-Caspian species in the Laurentian Great Lakes. Canadian Journal of Fish and Aquatic Science 59:1209–1228.

Vander Zanden, M.J., J.M. Casselman, and J.B. Rasmussen. 1999. Stable isotope evidence for the food web consequences of species invasions in lakes. Nature 401:464-467.

Vander Zanden, M.J., J.D. Olden, J.H. Thorne, and N.E. Mandrak. 2004. Predicting occurrences and impacts of smallmouth bass introductions in north temperate lakes. Ecological Applications 14:132-148.

Vehanen, T. and J. Aspi. 1996. Classification of northern Finnish lakes and the suitability for the stocking for brown trout (Salmo trutta (L.) m. lacustris). Fisheries Research 27:37-49.

Wagner, T., M.T. Bremigan, K.S. Cheruvelil, P.A. Soranno, N.A. Nate, and J.E. Breck. 2007. A multilevel modeling approach to assessing regional and local landscape features for lake classification and assessment of fish growth rates. Environmental Monitoring and Assessment 130:437-454.

Wehrly, K.E., J.E. Breck, L. Wang, L. Szabo-Kraft. 2012. A landscape-based classification of fish assemblages in sampled and unsampled lakes. Transactions of the American Fisheries Society 141:414-425.

Werner, R.G. 2004. Freshwater Fishes of the Northeastern United States. Syracuse University Press. Syracuse, NY.

Wiegmann D.D., J.R. Baylis, and M.H. Hoff. 1992. Sexual selection and fitness variation in a population of smallmouth bass, Micropterus dolomieui (Pisces: Centrarchidae). Evolution 46:1740-1753.

Wiegmann D.D., J.R. Baylis, and M.H. Hoff. 1997. Male fitness, body size and timing of reproduction in smallmouth bass, Micropterus Dolomieui. Ecology 78:111-128.

Wilde, G.R. 2003. Dispersal of tournament-caught black bass. Fisheries 28:10-17.

Winemiller, K.O. and K.A. Rose. 1992. Patterns of life-history diversification in

North American fishes: implications for population regulation. Canadian Journal of Fisheries and Aquatic Sciences 49:2196-2218.

Zipkin E.F., P.J. Sullivan, E.G. Cooch, C.E. Kraft, B.J. Shuter, and B.C. Weidel. 2008. Overcompensatory response of a smallmouth bass (Micropterus dolomieu) population to harvest: release from competition?

CHAPTER 1

Abstract

Geographically contiguous regionalization schemes have been used to effectively cluster waterbodies into useful subsets of shared biological and ecological characteristics. Doing so may provide valuable information about the biological and ecological status of unsampled waterbodies without the expense of time and labor-intensive field surveys. While the efficacy of geographically contiguous regionalization schemes in grouping waterbodies for fish assemblage characteristics has been repeatedly established, their utility in grouping waterbodies for fish demographic population metrics is less clear. Our purpose in this study was to determine the effectiveness of geographically contiguous regionalization schemes in grouping black bass population metrics. We found that lake-mean largemouth bass relative abundance was effectively clustered by the Ecological Drainage Unit (EDU) regionalization scheme and that lake-mean units of a metric of adult largemouth bass growth (length at age 5) were effectively clustered by both the EDU scheme and the four-digit Hydrologic Unit Code (HUC04) scheme. We suggest potential monitoring strategies for black bass demographic characteristics based on these results.

Introduction

Knowledge of community composition and population status is foundational to the effective management of sport fish populations. To obtain this knowledge for a specific water body, labor-intensive field surveys are typically required, and often the surveys must be repeated annually or on a regular schedule. However, in areas with abundant aquatic resources, it is impractical to conduct surveys on all waters on a frequent basis (Shuter et al. 1998). Therefore,

the ability to estimate community or population status by extrapolating what is known from a selection of well-monitored similar water bodies would be advantageous (Shuter et al. 1998, Irz et al. 2004, Tarkan et al. 2011, Cott et al. 2013). Strategies for classifying lakes have used both environmental parameters and geographic regionalization schemes (Table 1).

The more traditional strategy involves relating fish population or community response variables such as abundance or richness to environmental covariates, the nature of which may be internal to the waterbody (e.g. water chemistry, water temperature, flow velocity), morphological (e.g. surface area, depth), or external of the waterbody (e.g. ambient temperature, terrestrial catchment characteristics) (Table 1). This strategy has been used to explain variation in the fish community composition of unsampled lakes (Johnson et al. 1977, Tonn et al. 1983, Dolman 1990, and Schupp 1992) and to explain variation in the demographic characteristics of fish populations in unsampled lakes (Shuter et al 1998, Cott et al. 2013, Perry 2011). For example, Shuter et al. (1998) used measures of lake area and total dissolved solids to explain variation in life history characteristics of lake trout, *Salvelinus namaycush*. Similarly, Cott et al. (2013) explained variation in the life history characteristics of lake-dwelling burbot, *Lota lota*, using environmental covariates such as latitude, Secchi depth, lake area, and maximum lake depth. Perry (2011) used environmental variables collected from three spatial scales to explain variation in demographic variables of 11 warm-water fish species in hundreds of inland lakes in Indiana. While this strategy represents a cost and labor savings over conducting fisheries surveys in every water body, it may still require extensive field work to collect environmental data for waterbodies where no fisheries surveys have taken place. Although some environmental variables can be collected *en masse* using geographic information systems, other key variables such as water chemistry, temperature, depth and flow velocity require on site collections.

Despite management often being conducted on a waterbody-specific basis, management agency jurisdictions are often organized regionally. Therefore, a second strategy consisting of organizing water bodies based simply upon their shared membership in geographically contiguous regions may present a practical advantage over organizing them by their environmental attributes. Lake membership in geographically contiguous regions can be determined without site visits, potentially resulting in substantial time and labor savings. This lake-grouping strategy has been used to organize lakes according to their nutrient levels (Heiskary et al. 1987), water chemistry (Jenerette et al. 2002), water quality (Cheruvelil et al. 2008), and littoral macro-invertebrate assemblages (Johnson 2000). Geographically-contiguous regionalization schemes have also been used to organize streams by their benthic invertebrate assemblages (Pan et al. 2000, Rabeni and Doisy 2011, Moog et al. 2004), aquatic amphibian assemblages (Van Sickle and Hughes 2000), and notably, by their fish assemblage types (Larsen et al. 1986, Hughes et al. 1987, Hughes et al. 1990, Rohm et al. 1987, Newall and Magnuson 1999, Van Sickle and Hughes 2000, Oswood et al. 2000).

While the utility of using the regionalization strategy to predict fish community makeup in un-sampled water bodies has been repeatedly demonstrated (Larsen et al. 1986, Hughes et al. 1987, Hughes et al. 1990, Rohm et al. 1987, Newall and Magnuson 1999, Van Sickle and Hughes 2000), we are unaware of studies that suggest that this strategy is particularly effective in organizing water bodies by single-species population demographics. Wagner et al. (2007) evaluated the utility of two geographically-contiguous regionalization schemes (ecoregions and watersheds) in explaining variation in the growth of seven warm or cool-water fish species in Michigan lakes, but found that ecoregions were not effective and watersheds were only useful in organizing lakes for three out of 14 species-specific length-at-age analyses. They recommended

that other schemes besides ecoregions and watersheds be investigated.

In this study, we investigate the utility of using geographically-contiguous regionalization schemes to organize lakes in New York State by their species-specific black bass (largemouth bass, *Micropterus salmoides* and smallmouth bass *Micropterus dolomieu*) population characteristics. With over 7,850 lakes, New York State ranks sixth among U.S. states for number of lakes (behind Wisconsin, Maine, Michigan, Minnesota, and Alaska) (Anonymous 2009). As such, it is impractical for fisheries managers in New York to monitor black bass populations in most lakes, nor is it practical to maintain up to date records of water chemistry or other limnological variables that would presumably be subject to change over time for use in predictive modeling of fish population characteristics.

We explored the efficacy of seven different published regionalization schemes (Table 2) for organizing lakes by black bass growth (juvenile and adults considered separately), condition, and abundance. Since they are widespread in lakes throughout New York State and are an economically important sport fish for which much data are available in the state, black bass make a good test species for the efficacy of regionalization schemes. Our objective was to quantitatively rank the schemes according to their ability to cluster similar lakes in terms of black bass demographic characteristics and to determine which, if any, schemes were more effective than an empty model (a model with no schemes) in explaining variation in black bass population characteristics. We wanted to know which scheme minimized within-region variance while maximizing between-region variance in each of the demographic characteristics. We hypothesized that regionalization schemes delineated based upon environmental characteristics with either direct or indirect congruence with bass growth, relative condition, and relative abundance would outperform schemes lacking such relevance to bass population structure.

Methods

Fish Population Surveys

Largemouth bass and smallmouth bass were surveyed over a 24-year period from 1988-2011 by New York State Department of Environmental Conservation (NYSDEC) biologists in accordance with standardized sampling techniques (Green 1989). From these data, we summarized population metrics describing relative abundance (catch per unit effort), condition (relative weight), and growth of juveniles (length-at-age 2) and adults (length-at-age 5) for each species (Perry et al. 2014) (Table 3). Relative abundance metrics were derived exclusively from night boat electrofishing surveys. We only used samples where total amount of time spent electrofishing on a lake exceeded 30 minutes. Since condition and length-at-age metrics are not dependent on a specific gear type, in addition to using fish captured via boat electrofishing, condition and length-at-age metrics also utilized fish captured via trap net and gill net samples.

For all sample types, captured bass were identified, counted, and measured (total length to nearest mm). At least 5 fish per cm group were weighed to the nearest 5 g. Scales from at least 5 bass per cm group were collected for subsequent aging. In addition to these general survey methods, we further restricted our dataset for each of the population metrics to only those records that met specific criteria described below. In all, four population metrics were calculated for each species of bass and for both species, the number of lakes representing each population metric declined in the following order: CPUE, relative weight, length at age 2, and length at age 5 (Perry et al. 2014).

Relative Abundance

Prior to calculating relative abundance, we imposed an additional criterion that surveys

be conducted within a surface water temperature range of 15-25 degrees C, which corresponds with pre-spawning and spawning activities. To avoid biases associated with mixing juveniles and adults, we restricted our calculations of relative abundance to only adult fish. In the northern United States, largemouth bass mature around 3-4 years of age at an average length of 254-305mm (10-12 inches) (Mecozzi 1989) and smallmouth bass mature around 3-5 years at average lengths of 254-380mm (10-15 inches) (Wiegmann et al. 1992, 1997, Chu et al. 2006). Therefore, we excluded bass that were less than 254 mm in length from our calculations of adult abundance. To calculate abundance, the total number of adult fish was divided by the total hours of electrofishing conducted during the survey resulting in a catch per unit effort (CPUE in catch/hour) for each survey. Some lakes were only represented by one survey over the time period, but for lakes represented by >1 survey, we computed lake-mean CPUEs (mean of all survey CPUEs). This process resulted in single-survey CPUE's or lake-mean CPUEs of zero for some lakes. If the single-survey CPUE or lake-mean CPUE was zero, we checked all known survey records (regardless of gear type used) and if the species of bass had never been documented for that lake, we assumed that the species was not present in the lake and therefore eliminated the lake from further analysis of relative abundance for that species.

Relative Weight

As an indicator of the condition of black bass in lakes across the state, we summarized relative weights for largemouth bass and smallmouth bass at the survey and lake levels using species-specific standard equations (Henson 1991 for largemouth bass and Kolander et al. 1993 for smallmouth bass). Much of the variance in the body weight of an individual bass may be attributable to life history traits (such as reproductive state) that vary by season. Therefore, in an effort to reduce biases associated with seasonal variation, we restricted our set of records to only

those fish that were captured during the months of May or June. During these months of the year, a proportion of the weight of gravid females is attributed to the weight of the eggs. If sex ratios tended to vary across lakes, this could represent a source of bias where lakes with more females would tend to have higher relative weights. However the ratio of male to female black bass tends to be close to 50:50 in northern lakes (Beckman 1949). We only used those surveys that included 20 or more individual bass in our computations of relative weight.

Length-at-Age

We calculated bass length at age as an indicator of individual growth. For each species, specific age groups were selected to represent juvenile and adult bass separately (Perry et al. 2014). We then summarized this metric for each group by taking the mean length-at-age per survey. For lakes with only one survey, we used the single survey mean to represent the length at age for that lake; for lakes represented by more than one survey, we took the average of all survey means to represent the length at age. To reduce variation in length-at-age attributable to growth during the summer, we further restricted our dataset to only those fish caught during the spring season. Also, we only calculated the length-at-age metric for lakes that were represented by a minimum of 10 individual fish records (largemouth bass and smallmouth bass considered separately).

Regionalization Schemes

For this study, we compared the effectiveness of 7 published geographically-contiguous regionalization schemes (Table 2, Figure 1) with relevancy to lake bass populations in capturing among-region heterogeneity of four bass population metrics. These schemes were originally designed for purposes ranging from the examination of patterns and trends in natural resources, the management and conservation of natural resources, and the conservation of freshwater

biodiversity (Cheruvelil et al. 2008). All of these regionalization schemes have continuous coverage across the 48 contiguous continental United States, so would have wide applicability if found useful. Hereafter, we consistently refer to the subcomponent polygons of all seven regionalization schemes as 'regions'.

Prior to statistical analysis, surveys and lakes which contained relevant bass population metric data were mapped to the regions to which they belonged for each of the seven regionalization schemes (Table 2, Figure 1). For each of the bass population metrics, schemes with fewer, larger regions naturally had fewer 'empty' regions (regions not represented by data from at least one lake) and a greater number of lakes per region (Table 3, Table 4). For example, when the population metric was largemouth bass relative abundance, there were no regions in the ECO3 scheme that were not represented by at least 1 lake. In contrast, for that same population metric, there were 231 regions (68% of regions) in the HUC10 scheme that were not represented by a single lake (Table 3).

Statistical Analysis

For each black bass population metric (Table 3), our general strategy was to first determine whether any of the seven regionalization schemes (Table 2, Figure 1) were suitable for parsing out observed variability in the response metric and clustering lakes into useful ecological subsets. Additionally, if more than one regionalization scheme was found to be useful for a single bass population metric, we wanted to determine which scheme performed best. We used two different approaches to assess regionalization schemes for each population metric (Table 6). While both approaches relied on multi-level models, they differed primarily in whether the region variable was incorporated as a random or fixed effect.

The first approach, where region was included as a random variable, allowed us to

estimate the proportion of the variance explained at the region level and to use that information to help determine which regionalization schemes were useful. While this approach has been used by others in a similar context (Wagner et al. 2007), one drawback is that the region variable may not technically be considered to be random since regions were restricted to the geographic bounds of the state of New York. For example, for some metrics, every region in the scheme was occupied by a lake which was used in the model run. The result of this was that the region-level variance was subject to inflation.

The second approach, where region was included as a fixed variable, did not suffer from variance inflation, but the drawback to the second approach was that it was difficult to summarize the proportion of the total variance in the population metric that was attributable to among-region variance. Therefore, it was difficult to ascertain how well each scheme actually did at clustering lakes into useful biological or ecological subsets. The second approach was better suited to ranking the schemes to determine which scheme performed best in explaining between-region variance in the population metrics.

In both approaches, if model residuals did not meet the assumptions of having a normal distribution and equal variance, log-transformations of the response variable were applied and models were re-run. All models for both approaches were run in Program R (R Core Team 2014) using the lme4 package (Bates et al. 2014).

Approach #1: Region as a Random Effect.

Description of the model.

In this approach, 'region' was included as a random effect in a series of seven unconditional means multi-level models, one for each regionalization scheme considered. By definition, unconditional means models have no predictors included at any level.

Here, we used a two-level mixed-effect model (Equation 1) where the first level (Equation 2), representative of lakes, is nested within the second level (Equation 3), representative of regions. In the full two-level unconditional means model (Equation 1) *i* refers to lake and *j* refers to region. Y_{ij} is the lake-mean of the population metric. γ_{00} is the grand mean of the population metric across all regions, u_{0j} is a vector of the region-level residuals, and r_{ij} is a vector of the lake-level residuals.

Equation 1:

$$Y_{ij} = \gamma_{00} + u_{0j} + r_{ij}$$

The level 1 component of the unconditional means model is presented in Equation 2 where β_{0j} is the vector of region means. r_{ij} is a vector of lake-level (level 1) residuals resulting from subtracting the region mean, β_{0j} from each lake-mean, Y_{ij}. σ^2 is the variance of the lake-level (level 1) residuals (i.e., the within-region variability in the response metric).

Equation 2:

$$Y_{ij} = \beta_{0j} + r_{ij} \qquad r_{ij} \sim N(0, \sigma^2)$$

The level 2 component of the unconditional means model is presented in Equation 3 where u_{0j} is a vector of region-level (level 2) residuals resulting from subtracting the grand mean, γ_{00} from each region-mean, β_{0j}. τ_{00} is the variance of the region-level (level 2) residuals (i.e., the among-region variability in the response metric).

Equation 3:

$$\beta_{0j} = \gamma_{00} + u_{0j} \qquad u_{0j} \sim N(0, \tau_{00})$$

Since lakes were the unit of analysis, the input data consisted of the lake-means of survey-level data. For each population metric, a unique model was run for each of the seven regionalization schemes. The only component that differed across models was the particular

regionalization scheme that was included as a random effect. For this reason, restricted maximum likelihood (REML) was used instead of maximum likelihood (ML) to run each model, thereby justifying subsequent model comparisons (Zuur et al. 2009). For each population metric, the number of records (i.e. lakes) in the input data was the same for all seven models.

Identifying suitable regionalization schemes.

We considered that a regionalization scheme clustered lakes more effectively than random chance alone if the AIC score of the multi-level mixed-effect model which included that regionalization scheme was at least two units lower than the AIC score of an empty model – one with the same bass population response variable but with no regionalization scheme included (Akaike 1974, Burnham and Anderson 2002). Since the models being compared for a single population metric differed only in their random structure (in other words, the only difference between models was the regionalization scheme being used as the region-level random variable), restricted maximum likelihood (REML) was used (Zuur et al. 2009). For any given bass metric, if more than one schemes were associated with models with AIC scores at least two units lower than that of the empty model, we considered the scheme associated with the model with the lowest AIC score to be the most useful one, particularly if its AIC score was at least two units lower than the model for the next-best scheme. In this way the effectiveness of the schemes in clustering lakes for bass population metrics could be compared.

We also generated an F-statistic for each model by dividing the among-region variance (τ as defined above, or alternatively Mean Square Regression (MSR)) by the within-region variance (σ^2 as defined above, or alternatively Mean Square Error (MSE)). We ran spearman rank correlations between F-statistics and AIC scores, expecting that for each bass metric, the values of the model F-stats would be inversely related to the model AIC scores.

We also computed an intra-class correlation coefficient (ICC) for each model (Equation 4). The ICC has been used by others (Wagner et al. 2007, Cheruvelil et al. 2008) to evaluate the performance of clustering schemes in terms of the proportion of the total variance that is partitioned as region-level variance. We include ICC calculations here to improve comparability of our analysis with that of existing papers.

Equation 4:

$$ICC = \frac{\hat{\tau}_{00}}{(\hat{\tau}_{00} + \hat{\sigma}^2)}$$

We recognize that ICC is influenced by the degree of spatial autocorrelation relative to a scheme's region size (Figure 2). To account for the influence of spatial autocorrelation, we followed the general procedures used by Cheruvelil et al. (2008) for each population metric separately. First, we regressed the ICC against the mean region surface areas and the number of regions per scheme. Each regression consisted of seven points, one for each regionalization scheme. Second, we used regression tree analysis to find the optimal split of the response metric into two subsets with minimal within-subset residuals based on the schemes' mean region surface area and a separate regression tree analysis to identify the optimal split of the response metric based on the number of regions per scheme. Again, there were seven data points involved in each regression tree analysis, one point for each regionalization scheme. All regression trees were run in Program R (R Core Team 2014) using the rpart package (Therneau et al. 2014). Third, we generated a semi-variogram and estimated its range parameter. The range parameter describes the inter-lake distance (kilometers) beyond which no apparent spatial autocorrelation exists for the population metric. All semi-variograms were generated in Program R (R Core Team 2014) using the geoR package (Ribeiro and Diggle 2001).

Approach #2: Region as a Fixed Effect.

Description of the model.

In this approach, 'region' was incorporated as a fixed effect in seven unique models (one for each regionalization scheme) evaluated per bass population metric response variable. As with the first approach, here we used a two-level mixed-effect model; however, since as a fixed effect, region could not be incorporated as its own level in the model, we defined the levels differently. In this second approach, the first level (Equation 5), representative of surveys, is nested within the second level (Equation 6) which is representative of lakes. As a fixed-effect variable, 'region' is included at the lake level and describes the particular region to which each lake belongs.

The level 1 component of the model is presented in Equation 5 where β_{0i} is the vector of lake-mean values and r_{hi} is a vector of survey-level (level 1) residuals resulting from subtracting the lake-mean, β_{0i} from each survey value, Y_{hi}.

Equation 5:

$$Y_{hi} = \beta_{0i} + r_{hi}$$

The level 2 component of the model is presented in Equation 6 where u_{0i} is a vector of lake-level (level 2) residuals resulting from subtracting the grand mean, γ_{00} from each lake-mean, β_{0i}.

Equation 6:

$$\beta_{0i} = \gamma_{00} + \gamma_{0h}(\text{region}) + u_{0i}$$

As with the first approach, for each population metric, a unique model was run for each of the seven regionalization schemes.

Selecting the 'best' regionalization scheme.

We used this approach primarily as a means of effectively ranking all regionalization schemes and selecting the scheme that performed best. To do this, we again used AIC (Akaike, 1974) to compare models, only this time, we accounted for regions that were not represented by any lakes in our data by imposing an additional penalty consisting of two times the number of regions without lakes to the AIC score as a type of pseudo parameter equivalent in formulation to the standard penalty for estimated parameters. Our modified AIC equation is shown in Equation 7 where *L* is the maximized value of the likelihood function, k is the number of parameters in the model, and n is the number of regions not represented by black bass demographic data within the regionalization scheme.

Equation 7:

$$\boldsymbol{AIC = -2\ln(L) + 2k + 2n}$$

For each bass population metric, separate models were run for all seven regionalization schemes considered. In contrast to the first approach, here, all seven models associated with each bass population metric had the same random structure, and differed only in their fixed structure. For this reason, maximum likelihood (ML) was used instead of restricted maximum likelihood (REML) to run each model, thereby justifying subsequent model comparisons (Zuur et al. 2009).

Rank Correlation of Bass Metrics within Optimal Schemes

We also sought to determine whether bass population metrics were internally consistent within regionalization schemes that were considered optimal. For example, was a given region with relatively high abundance likely to have low condition and growth as expected given theoretical compensatory population dynamics? To investigate this notion, we ran spearman

rank correlations of population metrics averaged across the regions of the schemes that were considered optimal. Three schemes that were repeatedly identified as useful under approach #1 (when region was included in the models as a random variable) were the EDU, HUC04, and ECO3 schemes. Similarly, these three schemes were repeatedly identified as optimal under "Approach #2" (when region was included in models as a fixed variable). Given these results, we ran three separate intra-scheme correlations of population metrics for each bass species.

Results

Approach #1: Region as a Random Effect.

Generally, under the first approach, we found that for some population metrics (e.g. smallmouth bass relative abundance), none of the regionalization schemes were convincingly useful in explaining region-level variance, but for other population metrics (e.g. largemouth bass relative weight), one or more schemes were found to cluster lakes better than random chance alone. And in the case of the largemouth bass length at age 5 metric, we found that all seven of the schemes clustered lakes better than random chance alone.

Across all metrics and both species, the regionalization schemes which performed best (n = 5, listed in no particular order) were the HUC04 scheme (co-best for one metric), the HUC08 scheme (best for one metric), the HUC10 scheme (best for two metrics), the HEX scheme (co-best for one metric), and the EDU scheme (co-best for two metrics). While the ECO3 scheme clustered lakes better than random chance alone for the largemouth bass length at age 5 metric, other schemes performed significantly better. Similarly, although the ECO4 scheme clustered lakes better than chance alone for the largemouth bass relative abundance and length at age 5 metrics, other schemes performed significantly better in both cases. The ECO3, ECO4 and

HUC08 schemes never clustered lakes better than random chance alone for smallmouth bass population metrics.

Largemouth Bass

Overall, for the largemouth bass population metrics, we found that none of the seven regionalization schemes we evaluated clustered lakes better than random chance alone for the length at age 2 metric, only one scheme (HUC10) did so for the relative weight metric, six of the seven schemes did so for the relative abundance metric, and all seven schemes did so for the length at age 5 metric.

Relative Abundance (CPUE)

With the exception of the ECO3 scheme, all schemes clustered lakes better than random chance alone (Table 7). The model which included the HUC08 scheme had the lowest AIC score (22 units lower than that of the empty model) and that score was at least 2 AIC units lower (better) than the AIC scores of any other schemes' models. The models which included the HEX, ECO4 and EDU schemes had AIC scores which differed by less than two AIC units, but each of their AIC scores were at least two AIC units lower (better) than the AIC scores for the HUC04 or HUC10 schemes' models. Of the six models representing schemes that clustered lakes better than random chance alone, the model for the HUC10 scheme had the highest (worse) AIC score, yet that score was still 4.5 units lower (better) than that of the empty model. As expected, model F-statistics were negatively correlated with AIC scores (-0.81, p-value = 0.03). The HEX scheme had the highest ICC (34%), but five of the seven schemes had ICC's $\geq$ 10% and four schemes had ICC's greater than 20%. The median (±SD) ICC was 23% (±12%) (Table 7). The range parameter of the semi-variogram was 131.08 km which, if squared is 17,182 sq km (Figure 3). While the HUC04 scheme (16,752 sq km) and the ECO3 scheme (13,996 sq km)

had mean region surface areas very close to that, only the EDU scheme (31,636 sq km) had mean region surface areas greater than that.

Relative Weight

With an AIC score 2.3 units lower (better) than that of the empty model, only the HUC10 scheme clustered lakes better than random chance alone (Table 7). The correlation of model F-statistics and AIC scores was negative as expected (-0.86, p-value = 0.01). The HUC10 scheme also had the highest ICC, 34%; the ICC's of all other schemes being ≤ 10% (Table 7). The range parameter of the semi-variogram was 43.89 km which, if squared is 1,926 sq km (Figure 3). The HUC10 scheme (446 sq km) and the HEX scheme (649 sq km) were the only two schemes with mean region sizes that were less than that.

Length at Age 2

None of the seven schemes had delta AICs less than that of the empty model with its delta AIC of 0.0 (Table 7). Therefore, none of the schemes were determined to be useful in clustering lakes for this metric. As expected, the model F-statistics and AIC scores were negatively correlated (-0.96, p-value <0.01). Two of the seven schemes had ICC's ≥ 10%. The HUC10 scheme's ICC was 19% and the ECO3 scheme's ICC was 15%. All seven schemes had mean polygon sizes > 137 sq km, the square of the relatively low 11.71 km range parameter (Figure 3).

Length at Age 5

All seven schemes clustered lakes better than random chance alone (Table 7). The two models with the lowest (best) AIC scores (HEX and EDU schemes) had AIC scores which were less than two AIC units from each other, but had AIC scores which were at least two AIC units lower (better) than all other models. The sign of the correlation of the model F-statistics and

AIC scores was negative as expected (-0.25, p-value = 0.59). All seven schemes had ICC's ≥ 10% with the median (± SD) being a relatively high 45% (± 18%) (Table 7). The range parameter of the semi-variogram was 128.35 km which, if squared is 16,474 sq km (Figure 3). All but two schemes (EDU and HUC04) had mean region areas smaller than that.

Smallmouth Bass

For the smallmouth bass population metrics, we found that none of the schemes clustered lakes better than random chance alone for the relative abundance or relative weight metrics, only the HUC10 scheme did so for the length at age 2 metric, and three schemes (HUC04, EDU, and HEX) did so for the length at age 5 metric.

Relative Abundance (CPUE)

Although the AIC score of the model which included the HUC04 scheme was lower (better) than that of the empty model, the HUC04 scheme could not be considered to cluster lakes better than random chance alone because there was not sufficient separation in the AIC scores (the AIC score for the HUC04 scheme was not more than two AIC units lower than the empty model) (Table 8). The models of all other schemes had AIC scores which were higher (worse) than that of the empty model. As expected, model F-statistics were negatively correlated with AIC scores (-0.92, p-value < 0.01). The HUC10 scheme was the only scheme with an ICC ≥10% (Table 8). The range parameter of the semi-variogram was 179.98 km, which if squared is 32,392 sq km (Figure 4). All seven schemes had mean polygon sizes smaller than that (Table 8).

Relative Weight

None of the seven schemes had delta AICs less than that of the empty model with its delta AIC of 0.0 (Table 8). Therefore, none of the schemes were determined to be useful in clustering lakes for this metric. The model F-statistics and AIC scores were negatively

correlated as expected (-1.0, p-value <0.01). With its ICC of 46%, The HUC10 scheme was the only scheme with an ICC ≥10% (Table 8). The range parameter of the semi-variogram was 384.64 km, which if squared is 147,948 sq km (Figure 4). All seven of the schemes had mean polygon sizes smaller than that (Table 8).

Length at Age 2

With an AIC score 3.2 units lower (better) than that of the empty model, only the HUC10 scheme clustered lakes better than random chance alone (Table 8). The correlation of model F-statistics and AIC scores was negative as expected (-1.0, p-value < 0.01). Five of the seven schemes had ICC's ≥10% (Table 8). Indeed, across all seven schemes, the median (±SD) ICC was 24% (±32%) (Table 8). The HUC10 scheme had the highest ICC (84%). All but two schemes (EDU and HUC04) had mean polygon sizes smaller than the square (16,050 sq km) of the semi-variogram range parameter (126.69 km), (Table 8, Figure 4).

Length at Age 5

Three of the seven schemes (HUC04, EDU, and HEX) clustered the lakes better than random chance alone (Table 8). The model with the HUC04 scheme had the lowest delta AIC (0.0). That delta AIC was at least two AIC units lower (better) than that of the model with the HEX scheme (2.22), but it was less than two AIC units (1.98) from the delta AIC of the model with the EDU scheme. The sign of the correlation of the model F-statistics and AIC scores was negative as expected (-1.0, p-value < 0.01). Five of the seven schemes had models which explained ≥10% of the total variance at the region level (Table 8). Across all seven schemes, the median (±SD) ICC was 37% (±31%) and the ICC for the model with the hexagon scheme was a considerable 96% (Table 8). All schemes had mean region sizes smaller than the square (358,979sq km) of the semi-variogram range parameter (599.15 km), (Table 8, Figure 4).

An additional note about spatial autocorrelation:

For all four largemouth bass population metrics considered, we found that the ICC was negatively (although not significantly) related to mean region surface area (r^2 = .09 to 0.53; p-value ≤ 0.86; Figure 5) and positively (significantly only for CPUE and Wr) related to total number of regions (r^2 = 0.34 to 0.72; p-value ≤ 0.86; Figure 6).

For the smallmouth bass metrics, we also found that the ICC was negatively (although not significantly) related to mean region surface area (r^2 = .01 to 0.35; p-value ≤ 0.86; Figure 5) and for the length at age 2 and length at age 5 metrics, we found that the ICC was positively (although not significantly) related to total number of regions (length at age 2: r^2 = 0.01; p-value = 0.95; length at age 5: r^2 = 0.32; p-value = 0.18; Figure 6). For the relative abundance and relative weight metrics, we found that the ICC was negatively (although not significantly) related to total number of regions (relative abundance: r^2 = 0.04; p-value = 0.66; relative weight: r^2 = 0.06; p-value = 0.59; Figure 6).

Generally, these regressions indicate that spatial autocorrelation was an important factor in our analyses and that schemes which divide the landscape into a greater number of smaller regions were likely to explain a higher proportion of region-level variance simply because the lake-mean values of the population metric are auto-correlated (Figure 2).

As an example of the distribution of all four metrics for both species summarized by the regions of a single regionalization scheme, we present maps displaying this information when the scheme used is the EDU scheme (Figure 7, Figure 8). These maps help elucidate where differences in each individual metric occur among and between the regions of the EDU scheme. Moreover, the two sets of maps (one for each species) can be compared to identify spatial differences between species for each metric.

Approach #2: Region as a Fixed Effect.

By including region as a fixed effect, we were able to rank the seven schemes for each population metric and determine which scheme performed best in explaining among-region variance in each of the population metrics and the rankings were not confounded by potential variance inflation. As such, there was a top-ranked scheme (or, in the case of smallmouth bass length at age 2, a tie for top rank) for all of the eight (4 metrics x 2 species) population metrics evaluated. Overall, only three of the seven schemes were top-ranked or were tied for top-ranking across all eight population metrics (Table 9). The scheme that was most often ranked highest was the HUC04 scheme which was the top-ranking scheme for three of the eight metrics and also tied for top rank for a fourth metric (Table 9). The ECO3 scheme was the top ranked for three population metrics and also tied for top rank for a fourth metric. Finally, the EDU scheme was the top ranked for one population metric.

Largemouth Bass

Relative Abundance (CPUE)

For largemouth bass CPUE, the scheme with the lowest penalized AIC score was the EDU scheme (Table 10). Since none of the other schemes had delta AIC values within two units of the penalized AIC score for the EDU scheme, EDU was identified as the optimal scheme for explaining among-region variance for largemouth bass CPUE.

Relative Weight

For largemouth bass relative weight, the HUC04 scheme had the lowest penalized AIC score and was at least two AIC units lower (better) than all other schemes (Table 10). Therefore, the HUC04 scheme ranked as the best of the seven schemes evaluated in explaining among-region variance for largemouth bass relative weight.

Length at Age 2

For the largemouth bass juvenile growth metric, length at age 2, the ECO3 scheme had the lowest penalized AIC score which was at least two AIC units lower than all other schemes (Table 10), making it the best of the seven schemes evaluated in explaining among-region variance for largemouth bass length at age 2.

Length at Age 5

For the largemouth bass adult growth metric, length at ages 5, the ECO3 scheme had the lowest penalized AIC score and that AIC score was at least two AIC units lower than all other schemes (Table 10), making it the best of the seven schemes evaluated in explaining among-region variance for largemouth bass length at age 5.

Smallmouth Bass

Relative Abundance (CPUE)

For smallmouth bass CPUE, the scheme with the lowest penalized AIC score was the HUC04 scheme (Table 10). Since none of the other schemes had delta AIC values within two units of the penalized AIC score for the HUC04 scheme, HUC04 was identified as the optimal scheme for explaining among-region variance for smallmouth bass CPUE.

Relative Weight

For smallmouth bass relative weight, the ECO3 scheme had the lowest penalized AIC score and that AIC score was at least two AIC units lower (better) than all other schemes (Table 10). Therefore, the ECO3 scheme ranked as the best of the seven schemes evaluated in explaining among-region variance for smallmouth bass relative weight.

Length at Age 2

For the smallmouth bass juvenile growth metric, length at age 2, the ECO3 scheme had

the lowest penalized AIC score (delta AIC = 0), but that score was less than two AIC units away from that of the HUC04 scheme (delta AIC = 0.1) (Table 10). Therefore, both the ECO3 and HUC04 schemes were considered to be jointly optimal of the seven schemes evaluated in explaining among-region variance for smallmouth bass length at age 2.

Length at Age 5

For the smallmouth bass adult growth metric, length at ages 5, the HUC04 scheme had the lowest penalized AIC score and that AIC score was at least two AIC units less (better) than all other schemes (Table 10), making it the best of the seven schemes evaluated in explaining among-region variance for smallmouth bass length at age 5.

Rank Correlation of Bass Metrics within Optimal Schemes

We found that the spearman rank correlations provided mixed results in terms of demonstrating the expected pattern (Table 11). For example, a negative correlation would be expected between relative abundance and condition (relative weight), but in four out of six such correlations, the relationship was positive. The sign of all inter-metric correlations were as expected for smallmouth bass when the HUC04 scheme was used, but two of the six correlations for smallmouth bass in each of the other two schemes had signs that were counter-intuitive and for largemouth bass, the signs were counter-intuitive for two to four of the six correlations depending on which scheme was used. In total, only 64% of correlations had the expected sign.

Discussion

Our results suggest that regionalization schemes may be relied upon to cluster lakes better than random chance alone for the four population metrics we considered across two black bass species. We also found that some schemes perform better than others in clustering lakes for

bass metrics. The proportion of the model variance explained at the region level, i.e., the ICC, ranged from 9% to 96% depending on the metric and regionalization scheme considered. These percentages were generally higher than those found in a similar study evaluating the usefulness of regionalization schemes in grouping lakes by fish growth rates in the state of Michigan (Wagner et al. 2007).

Despite these findings, we caution against the use of regionalization schemes as the primary tools for perpetual monitoring of state-wide black bass population metrics in lakes. It is important to recognize that even in cases where quantitative criteria are satisfied, a scheme still may not cluster lakes in a way consistent with biological or ecological differences. Citing studies involving algae, aquatic invertebrates, fish and other aquatic vertebrates in both streams and lakes, Hawkins et al. (2000) noted that even though landscape classifications may account for more biotic variation than would be expected by chance alone, they have only limited use in aquatic ecosystem management (Hawkins et al. 2000).

If a regionalization scheme was effective in partitioning lacustrine bass population metrics, correlations of metrics summarized for each region should show consistent patterns. For example, region-summarized bass growth metrics should show positive correlations and, due to compensatory population dynamics, the correlation between relative abundance and growth summarized at the region level would be expected to be negative. However, when we ran rank correlations of all four bass metrics for each species using three schemes that performed well, the results were rather mixed.

These results suggest that while the use of regionalization schemes as the primary tool for devising long-term state-wide lake monitoring plans for bass population metrics is perhaps ill-advised, regionalization schemes may still have merit and serve an important purpose for

managers seeking to understand regional patterns of black bass population characteristics. Hawkins et al. (2000) suggest that despite their limited use in aquatic bioassesments, regionalization schemes may still play an important role in aquatic management in that they "…provide an initial stratification of site locations to ensure that different landscape features are adequately represented in a sampling program." We agree and suggest that regionalization schemes may be relied upon to identify specific regions that deviate from broader statewide patterns in terms of black bass population metrics.

The appeal of regionalization schemes as a method of clustering lacustrine bass populations relies, to some extent, upon the notion that bass population dynamics are subject to strong geographic influences and that environmental characteristics influential in driving lacustrine bass populations can be organized at relatively large spatial scales and can be delineated regionally. An example of such an environmental characteristic is topographic elevation. In New York State, elevation varies from sea level in New York City, Long Island, and coastal areas of the Laurentian Great Lakes to over 900 meters (3,000 feet) in the Adirondack and Catskill mountains. Lakes at higher elevations would be expected to have cooler temperatures, be fed by lower order streams, and have less nutrient loading and as a result, bass population metrics such as abundance, growth and condition might be expected to be generally lower in higher elevation lakes than in lower elevation lakes. In this study, we found that the EDU scheme explained significant percentages of among-region variance in the largemouth bass relative abundance metric and that the positioning of the regions with the most extreme differences in mean relative abundance within that scheme was consistent with expectations from elevational differences.

The somewhat lackluster degree to which regionalization schemes partitioned variance in

black bass population metrics in this study suggests either that influential environmental characteristics were not adequately segregated within the schemes that we chose to evaluate or that environmental characteristics which cannot be easily delineated at a large spatial scale are more influential in driving bass population characteristics. Of course, the role of biological interactions such as competition and predation in driving bass population metrics should be recognized as well.

The strategy we used in our second approach of imposing an additional penalty to the AIC score of each model representing the different regionalization schemes in order to account for unused regions is an innovative aspect of this study. By multiplying the number of regions from which no data were available times 2 and adding that product to the AIC score, we were able to include lake as a random variable and our results became applicable to all regions within the scheme rather than just the regions containing sampled lakes. A similar AIC-penalizing strategy could be used in the ranking of regionalization schemes in order to incorporate physical or labor costs associated with sampling efforts, qualitative valuations of specific lakes, or environmental concerns.

In chapter two, we investigate the role of eleven different environmental predictor variables (e.g. lake depth, elevation, and mean summer air temperature) summarized at survey, lake and region levels in improving the explanatory power of the models we've described here. In doing so, we identify which environmental variables have the most influence in driving bass population metrics.

References

Akaike, H. (1974). A new look at the statistical model identification. IEEE Transactions on

Automatic Control 19: 716-723.

Albert, D.A. (1995). Regional landscape ecosystems of Michigan, Minnesota, and Wisconsin: A working map and classification. USDA Forest Service North Central Forest Experiment Stateion General Technical Report (NC-178). Retrieved from (http://www.treesearch.fs.fed.us/pubs/10242).

Anonymous (2009). Diet of a small lake. New York Department of Environmental Convservation. 313 pp. Retrieved from (http://www.dec.ny.gov/chemical/82123.html).

Bates, D., M. Maechler, B. Bolker and S. Walker (2014). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.0-5. http://CRAN.R-project.org/package=lme4.

Brown, R.S., and K. Marshall (1996). Ecosystem management in state governments. Ecological Applications, 6, 721–723.

Burnham, K. P., & Anderson, D. R. (2002). Model selection and multimodel inference: a practical information-theoretic approach. Springer Science & Business Media.

Cheruvelil, K.S., P.A. Soranno, M.T. Bremigan, T. Wagner, S.L. Martin (2008). Grouping lakes for water quality assessment and monitoring: the roles of regionalization and spatial scale. Environmental Management, 41, 425-440.

Connelly, N.A., T.L. Brown, and B.A. Knuth (1997). New York Statewide Angler Survey 1996, Report 1: Angler effort and expenditures. New York State Department of Environmental Conservation, Bureau of Fisheries, Albany, NY.

Connelly N.A., and T.L. Brown. (2009). New York Statewide Angler Survey, 2007. New York State Department of Environmental Conservation, Bureau of Fisheries, Albany, NY.

Cott, P.A., T.A. Johnston, and J.M. Gunn. Stability in life history characteristics among burbot populations across environmental gradients. Transactions of the American Fisheries Society, 142, 1746-1756.

Dolman, W.B. (1990). Classification of Texas reservoirs in relation to limnology and fish community associations. Transactions of the American Fisheries Society, 119, 511-520.

Duncan, G. J., and G. Kalton. (1987). Issues of design and analysis of surveys across time. International Statistical Review, 55, 97–117.

Green, D.M. (1989). Centrarchid Sampling Manual. New York State Department of Environmental Conservation, Bureau of Fisheries, Albany, NY.

Hawkins, C. P., Norris, R. H., Gerritsen, J., Hughes, R. M., Jackson, S. K., Johnson, R. K., & Stevenson, R. J. (2000). Evaluation of the use of landscape classifications for the prediction of freshwater biota: synthesis and recommendations. Evaluation, 19(3), 541-556.

Heiskary, S.A., C.B. Wilson, D.P. Larsen (1987). Analysis of regional patterns in lake water quality: using ecoregions for lake management in Minnesota. Lake and Reservoir Management, 3, 337–344.

Henson, J.C. (1991). Quantitative description and development of a species-specific growth form for largemouth bass, with application to the relative weight index. Master's thesis. Texas A&M University, College Station.

Higgins J.V., M.T. Bryer, M.L. Khoury, T.W. Fitzhugh (2005). A freshwater classification approach for biodiversity conservation planning. Conservation Biology, 19, 432-445.

Hughes, R.M., E. Rexstad, andC.E. Bond (1987). The relationship of aquatic ecoregions, river catchments and physiographic provinces to the ichthyogeographic regions of Oregon. Copeia, 1987, 423-432.

Hughes, R.M., T.R. Whittier, and C.M. Rohm (1990). A regional framework for establishing recovery criteria. Environmental Management, 14, 673-683.

Irz, P., C. Argillier, and T. Oberdorff (2004). Native and introduced fish species richness in French lakes: local and regional influences. Global Ecology and Biogeography, 13, 335-344.

Jenerette G.D., J. Lee, D.W. Waller, R.E. Carlson (2002). Multivariate analysis of the ecoregion delineation for aquatic systems. Environmental Management, 29, 67–75.

Johnson, M.G., J.H. Leach, C.K. Minns, and C.H. Olver (1977). Limnological characteristics of Ontario lakes in relation to associations of walleye (Stizostedion vitreum vitreum), northern pike (Esoc lucius), lake trout (Salvelinus naymaycush), and smallmouth bass (Micropterus dolomieui). Journal of the Fisheries Board of Canada, 34, 1592-1601.

Johnson R.K. (2000). Spatial congruence between ecoregions and littoral macroinvertebrate assemblages. Journal of the North American Benthological Society, 19, 75–486.

Kolander, T.D., D.W. Willis, and B.R. Murphy (1993). Proposed revision of the standard weight (Ws) equation for smallmouth bass. North American Journal of Fish Management 13: 398-400.

Larsen, D.L., J.M. Omernik, R.M. Hughes, C.M. Rohm, T.R. Whittier, A.J. Kinney, A.L. Gallant, and D.R. Dudley (1986). Correspondence between spatial patterns in fish assemblages in Ohio streams and aquatic ecoregions. Environmental Management 10, 815-828.

Lazarsfeld, P., and M. Fiske (1938). The" panel" as a new tool for measuring opinion. The Public Opinion Quarterly, 2(4), 596-612.

Maunder M.N., and A.E. Punt (2004). Standardizing catch and effort data: a review of recent approaches. Fisheries Research 70: 141-159.

Mecozzi M. (1989). Largemouth Bass Fact Sheet. Wisconsin Department of Natural Resources Bureau of Fisheries Management PUBL-FM-709 89

Moog O., A. Schmidt-Kloiber, T. Ofenbock, J. Gerritsen (2004). Does the ecoregion approach support the typological demands of the EU Water Framework Directive? Hydrobiologia, 516, 21–23.

Newall P.R., J.J. Magnuson (1999). The importance of ecoregion versus drainage area on fish distributions in the St. Croix River and its Wisconsin tributaries. Environmental Biology of Fishes, 55, 245–254.

Omernik, J.M. (1987). Ecoregions of the conterminous United States. Annals of the Association of American Geographers, 77, 118-125.

Omernik, J.M. (1995). Ecoregions: A spatial framework for environmental management. In biological assessment and criteria: Tools for water resource planning and decision making, W.S. Davis and T.P. Simon, eds. Lewis Publishers, Boca Raton, FL, 49-62.

Oswood, M. W., Reynolds, J. B., Irons III, J. G., & Milner, A. M. (2000). Distributions of freshwater fishes in ecoregions and hydroregions of Alaska*. Journal of the North American Benthological Society, 19(3), 405-418.

Pan Y., R.J. Stevenson, B.H. Hill, A.T. Herlihy (2000). Ecoregions and benthic diatom assemblages in the Mid-Atlantic Highlands streams, USA. Journal of the North American Benthological Society, 19, 518–540.

Perry, P.C. (2011). The role of compensatory dynamics and influence of environmental factors across multiple spatial scales in structuring fish populations. Masters Thesis. Purdue University department of forestry and natural resources.

Perry, P.C., J.J. Loukmas, W.L. Fisher, P.J. Sullivan, and J.R. Jackson. (2014). Characterizing the status of black bass populations in New York. Final Report. New York State Department of Environmental Conservation, Albany, New York.

Peterson, J.T. and T.J. Kwak. (1999). Modeling the effects of land use and climate change on riverine smallmouth bass. Ecological Applications, 9, 1391-1404.

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Rabeni, C.F. and K.E. Doisy (2000). Correspondence of stream benthic invertebrate assemblages to regional classification schemes in Missouri. Journal of the North American Benthological Society, 19, 419-428.

Ribeiro Jr., P.J. and P.J. Diggle (2001). geoR: a package for geostatistical analysis. R-NEWS, 1, 15-18.

Rohm, C.M., J.W. Giese, and C.C. Bennett (1987) Evaluation of aquatic ecoregion classification of streams in Arkansas. Journal of Freshwater Ecology, 4, 127-140.

Schupp, D.H. (1992). An ecological classification of Minnesota lakes with associated fish communities. Minnesota Department of Natural Resources Investigational Report, 417, 1-27.

Scott J.M., F. Davis, B. Csuti, R. Noss, B. Butterfield, C. Groves, H. Anderson, S. Caicco, F. D'Erchia, T.C. Edwards, J. Ulliman, R.G. Wright (1993.) Gap Analysis: A Geographic Approach to Protection of Biological Diversity. Wildlife Monographs, 123, 3-41

Scott J.M., T.H. Tear, F. Davis (1996.) Gap Analysis: A landscape approach to biodiversity planning. American Society for Photogrammetry and Remote Sensing. Bethesda, Maryland.

Seaber, P. R., F.P. Kapinos, and G.L. Knapp (1987). Hydrologic Unit Map, USGS Water-Supply Paper 2294.

Shuter, B.J., M.L. Jones, R.M. Korver, and N.P. Lester (1998). A general, life history based model for regional management of fish stocks: the inland lake trout (Salvelinus namaycush) fisheries of Ontario. Canadian Journal of Fisheries and Aquatic Science, 55, 2161-2177.

Tarkan, A.S., Ö. Gaygusuz, M.J. Godard, G.H. Copp (2011). Long-term growth patterns in a pond-dwelling population of crucian carp, Carassius carassius: environmental and density-realted factors. Fisheries Management and Ecology, 18, 375-383.

Therneau, T., B. Atkinson, and B. Ripley (2014). rpart: Recursive Partitioniing and Regression Trees. R package version 4.1.8. http://CRAN.R-project.org/package=rpart

Tillma, J.S., C.S. Guy, and C.S. Mammoliti. (1998). Relations among habitat and population characteristics of spotted bass in Kansas streams. North American Journal of Fisheries Management, 18, 886-893.

Tonn, W.M., J.J. Magnuson, A.M. Forbes (1983). Community analysis in fishery management: an application with northern Wisconsin lakes. Transactions of the American Fisheries Society, 112, 368-377.

Urquhart, N.S., S. G. Paulsen, and D. P. Larsen, (1998). Monitoring for policy-relevant regional trends over time. Ecological Applications, 8, 246-257.

Vanni M.J., K.K. Arend, M.T. Bremigan, D.B. Bunnell, J.E. Garvey, M.J. Gonzalez, W.H.

Renwick, P.A. Soarnno, and R. A. Stein (2005). Linking landscapes and food webs: effects of omnivorous fish and watersheds on reservoir ecosystems. Bioscience 55, 155-167.

Van Sickle, J., and R.M. Hughes (2000). Classification strengths of ecoregions, catchments, and geographic clusters for aquatic vertebrates in Oregon. Journal of the North American Benthological Society, 19, 370–384.

Wagner T., M.T. Bremigan, K.S. Cheruvelil, P.A. Soranno, N.A. Nate, J.E. Breck (2007). A multilevel modeling approach to assessing regional and local landscape features for lake classification and assessment of fish growth rates. Environmental Monitoring and Assessment, 130, 437–454.

Wege, G.J. and R.O. Anderson (1978). Relative weight (Wr): a new index of condition for largemouth bass. Pages 79-91 in G.D. Novinger and J.G. Dillard, editors. New approaches to the management of small impoundments. American Fisheries Society, North Central Division, Special Publication 5, Bethesda, Maryland.

Zuur, A.F., E.N. Ieno, N.J. Walker, A.A. Saveliev, and G.M. Smith (2009). Mixed effects models and extensions in ecology with R. Springer, New York, NY. pp 116-122.

Tables

Table 1: Comparison of two strategies used to extend fisheries community and population information from surveyed water bodies to un-surveyed water bodies

Strategy	Description	Community Makeup Examples	Single-Species Demographic Examples	Example covariates that can be collected *en masse* using GIS	Example of covariates that require a site visit	Time and Labor Requirements
1	Relate environmental covariates to a fish community or single-species demographic response variable.	Johnson et al. 1977, Tonn et al. 1983, Dolman 1990, Schupp 1992	Shuter et al. 1998, Perry 2011, Cott et al. 2013	lake surface area, perimeter, river length, land cover and land use characteristics	water chemistry, maximum lake depth, mean lake depth, stream velocity	Potentially high
2	Organize fish community or single-species demographic characteristic based upon the shared regional membership of water bodies.	Larsen et al. 1986, Hughes et al. 1990, Rohm et al. 1987, Newall and Magnuson 1999, Van Sickle and Hughes 2000	Wagner et al. 2007, the present study	water body membership in geographically-contiguous regionalization schemes	N/A	Low

Table 2: Description of the seven published regionalization schemes used in this study. The efficacy of each scheme in explaining among-region variance in multiple black bass population metrics was evaluated. Refer to Figure 1 for maps of each scheme. The number and size summaries of regions refer only to regions within New York State.

Regionalization Scheme	Abbrev. Name	Description	Published Source	#Regions in New York	Min size (km^2)	Max size (km^2)	Median Size (km^2)	Mean Size (km^2)	SD
Omernik's Ecoregion, version III	**ECO 3**	Delineated based on observed patterns in land use, land surface form, natural vegetation, and soils.	Omernik 1987	9	514	36,132	3,086	13,996	16,320
Omernik's Ecoregion, version IV	**ECO 4**	A sub-delineation of Omernik's Ecoregion, version III.	Omernik 1995	42	27	20,550	1,681	2,999	3,994

Regionalization Scheme	Abbrev. Name	Description	Published Source	#Regions in New York	Min size (km^2)	Max size (km^2)	Median Size (km^2)	Mean Size (km^2)	SD
Ecological Drainage Unit	**EDU**	Delineated to conform to patterns of physiography, climate, and freshwater ecosystem connectivity (i.e., the networks formed by freshwater systems, including lakes, wetlands, glaciers, streams, and coastal waters).	Higgins et al. 2005	12	6,681	51,007	31,677	31,636	14,317
US EPA EMAP equal-area hexagons	**HEX**	Environmental monitoring and assessment program (EMAP) grids used for regular interval biological sampling.	Scott et al. 1993, 1996	249	648	648	648	648	-
USGS 4-digit Hydrologic Unit	**HUC04**	The "subregion" level of the USGS hydrologic classification system of surface water drainage areas.	Seaber et al. 1987	10	6,212	37,273	14,101	16,752	10,625
USGS 8-digit Hydrologic Unit	**HUC08**	The "subbasin" level of the USGS hydrologic classification system of surface water drainage areas.	Seaber et al. 1987	55	376	8,959	2,718	3,046	1,823
USGS 10-digit Hydrologic Unit	**HUC10**	The "watershed" level of the USGS hydrologic classification system of surface water drainage areas.	Seaber et al. 1987	339	66	3,081	406	446	406

Table 3: Summary of the lakes for which we had largemouth bass population metric data for each of the seven regionalization schemes

Bass pop. metric:	Region. scheme:	Total # of regions:	Regions with lakes:	Regions without lakes:	% without lakes:	Mean # of lakes per region:	SD:	Total # of lakes represented:
CPUE	ECO 3	9	9	0	0%	22.78	18.40	
	ECO 4	42	33	9	21%	6.21	5.91	
	EDU	12	12	0	0%	17.08	14.31	
	HEX	249	90	159	64%	4.51	21.36	**205**
	HUC04	10	10	0	0%	20.50	22.26	
	HUC08	55	41	14	25%	5.00	6.00	
	HUC10	339	108	231	68%	1.90	1.71	
Relative Weight	ECO 3	9	9	0	0%	16.22	13.5	
	ECO 4	42	29	13	31%	5.03	5.11	
	EDU	12	12	0	0%	12.17	11.67	
	HEX	249	73	176	71%	2	1.91	**75**
	HUC04	10	10	0	0%	14.6	17.24	
	HUC08	55	37	18	33%	3.95	4.95	
	HUC10	339	81	258	76%	1.8	1.54	
Length at Age 2	ECO 3	9	7	2	22%	7.43	7.44	
	ECO 4	42	19	23	55%	2.74	3.89	
	EDU	12	10	2	17%	5.2	6.6	
	HEX	249	33	216	87%	1.58	1.71	**52**
	HUC04	10	9	1	10%	5.78	5.61	
	HUC08	55	23	32	58%	2.26	2.49	
	HUC10	339	35	304	90%	1.45	1.27	
Length at Age 5	ECO 3	9	7	2	22%	6.71	5.82	
	ECO 4	42	15	27	64%	3.13	3.6	
	EDU	12	8	4	33%	5.88	5.54	
	HEX	249	28	221	89%	1.68	1.49	**47**
	HUC04	10	8	2	20%	5.88	3.68	
	HUC08	55	21	34	62%	2.24	1.97	
	HUC10	339	30	309	91%	1.57	1.01	

Table 4: Summary of the lakes for which we had smallmouth bass population metric data for each of the seven regionalization schemes

Bass Pop. Metric:	Region. Scheme:	Total # of regions:	Regions with lakes:	Regions without lakes:	% without lakes:	Mean # of lakes per region:	SD:	Total # of lakes represented:
CPUE	ECO 3	9	9	0	0%	14.44	14.30	
	ECO 4	42	27	15	36%	4.81	3.20	
	EDU	12	11	1	8%	11.82	7.55	
	HEX	249	72	177	71%	1.81	1.24	**130**
	HUC04	10	9	1	10%	14.44	10.00	
	HUC08	55	38	17	31%	3.42	2.76	
	HUC10	339	82	257	76%	1.59	1.08	
Relative Weight	ECO 3	9	7	2	22%	10.57	9.31	
	ECO 4	42	24	18	43%	3.08	1.93	
	EDU	12	10	2	17%	7.4	4.24	
	HEX	249	52	197	79%	1.42	0.75	**74**
	HUC04	10	8	2	20%	9.25	5.75	
	HUC08	55	33	22	40%	2.24	1.44	
	HUC10	339	56	283	83%	1.32	0.83	
Length at Age 2	ECO 3	9	6	3	33%	5.33	4.18	
	ECO 4	42	18	24	57%	1.78	1.11	
	EDU	12	9	3	25%	3.56	1.59	
	HEX	249	25	224	90%	1.28	0.74	**32**
	HUC04	10	8	2	20%	4	1.41	
	HUC08	55	21	34	62%	1.52	0.68	
	HUC10	339	27	312	92%	1.19	0.56	
Length at Age 5	ECO 3	9	6	3	33%	4	2.61	
	ECO 4	42	15	27	64%	1.6	0.74	
	EDU	12	9	3	25%	2.67	1.12	
	HEX	249	22	227	91%	1.09	0.29	**24**
	HUC04	10	8	2	20%	3	0.93	
	HUC08	55	19	36	65%	1.26	0.56	
	HUC10	339	21	318	94%	1.14	0.48	

Table 5: Black bass population metrics. Statistical summaries are provided for lake-mean values.

	Largemouth Bass						Smallmouth Bass					
	#Lakes	Min	Max	Median	Mean	SD	#Lakes	Min	Max	Median	Mean	SD
CPUE	205	0	114	11	17	19	130	0	45	4	7	9
Wr	146	83	123	99	99	7	74	75	120	88	90	8
Length at Age 2 (mm)	52	120	262	182	185	28	32	125	265	168	173	32
Length at Age 5 (mm)	47	225	405	329	332	35	24	225	373	319	319	35

Table 6: Comparison of the two statistical approaches used to evaluate the utility of each regionalization scheme in explaining variance in black bass population metrics.

	Approach 1	Approach 2
Number of regionalization schemes considered:	7	7
Region incorporated as random or fixed:	Random	Fixed
Form of maximum likelihood estimation used:	REML	ML
Model Level 1: Units of analysis:	Lake-means of surveys	Surveys
Model level 2: Random variable:	Regions	Lakes
Bass population metrics considered:	CPUE, Wr, length at age 2, length at age 5	CPUE, Wr, length at age 2, length at age 5
Total number of model runs:	64. [2 species x 4 pop. metrics x (7 schemes + 1 empty model)]	56. [2 species x 4 population metrics x 7 schemes]
A benefit of the approach:	The proportion of region-level variance explained in the response variable by each regionalization scheme can be determined as well as whether or not this represents a significant improvement over an "empty" model with no schemes.	Technically, region should be considered a fixed effect due to the fact that the number of regions is limited by the geographic boundaries of New York State. This approach correctly recognizes that the selection of regions to include in a model is not truly random.

	Approach 1	Approach 2
A drawback of the approach:	The potential exists for region-level variance inflation to occur as a result of considering region to be a random effect when in fact it is constrained to the state of New York for all regionalization schemes.	It is difficult to determine the proportion of the total variance explained at the region level; therefore, it is difficult to assess how well each scheme actually did at clustering lakes into useful biological or ecological subsets.
How to determine if models with regionalization schemes performed better in explaining variance in the response than "empty" models which did not include regionalization schemes:	AIC model comparison was used to identify whether each regionalization scheme clustered lakes better than random chance alone for each bass metric.	No clear way to determine (see drawback).
How "best" scheme was selected:	The "best" scheme was chosen by selecting the scheme associated with the model that had the lowest AIC score. For each response metric, models which included regionalization schemes were only considered if they were first deemed better than an "empty" model.	A penalized-AIC score was generated for each model by incorporating the number of un-occupied regions for each scheme, and then the 7 schemes were ranked based on their penalized-AIC scores.

Table 7: Largemouth bass multi-level model results (where region was assigned as a random variable) to determine which of seven regionalization schemes were useful in explaining region-level variance in four different population metrics. For metrics where one or more scheme was determined to be useful, gray shading indicates the 'best' model.

Bass pop. metric:	Multi-level model:	Total variance	Lake-level variance	Region-level variance	ICC	F-STAT	Δ AIC	Mean region area (km^2)	Passes AIC criterion
CPUE	(EMPTY)						22.0		
	ECO 3	1.28	1.28	0.00	0.00	0.00	24.0	13996	
	ECO 4	1.32	1.01	0.31	0.23	0.31	6.1	2999	✓
	EDU	1.28	1.11	0.18	0.14	0.16	7.4	31636	✓
	HEX	1.31	0.86	0.45	0.34	0.52	6.0	649	✓
	HUC04	1.29	1.18	0.11	0.09	0.10	15.5	16752	✓
	HUC08	1.33	0.95	0.38	0.29	0.40	0.0	3046	✓
	HUC10	1.29	0.99	0.30	0.23	0.31	17.5	446	✓
Wr	(EMPTY)						2.3		
	ECO 3	43.2	41.78	1.42	0.03	0.03	2.9	13996	
	ECO 4	43.1	42.06	1.04	0.02	0.02	3.9	2999	
	EDU	43.7	41.16	2.55	0.06	0.06	2.8	31636	
	HEX	43.11	38.74	4.37	0.10	0.11	2.6	649	
	HUC04	44.45	40.92	3.54	0.08	0.09	2.5	16752	
	HUC08	43.33	41.72	1.61	0.04	0.04	4.2	3046	
	HUC10	45.6	30.05	15.55	0.34	0.52	0.0	446	✓
Length at Age 2	(EMPTY)						0.0		
	ECO 3	844.5	721.60	122.90	0.15	0.17	0.3	13996	
	ECO 4	807.8	807.80	0.00	0.00	0.00	2.0	2999	
	EDU	807.8	807.80	0.00	0.00	0.00	2.0	31636	
	HEX	811.37	769.15	42.22	0.05	0.05	2.0	649	
	HUC04	807.8	807.80	0.00	0.00	0.00	2.0	16752	
	HUC08	807.8	807.80	0.00	0.00	0.00	2.0	3046	
	HUC10	826.2	666.70	159.50	0.19	0.24	1.8	446	
Length at Age 5	(EMPTY)						9.1		
	ECO 3	1670.1	812.40	857.70	0.51	0.36	2.8	13996	✓
	ECO 4	1413.8	787.10	626.70	0.44	0.83	5.0	2999	✓
	EDU	1192.2	844.20	348.00	0.29	1.77	1.1	31636	✓
	HEX	1536	345.30	1190.70	0.78	3.45	0.0	649	✓
	HUC04	1252.1	919.00	333.10	0.27	0.41	5.0	16752	✓
	HUC08	1268.9	693.20	575.70	0.45	1.06	3.2	3046	✓
	HUC10	1380.2	497.70	882.50	0.64	0.80	5.6	446	✓

Table 8: Smallmouth bass multi-level model results (where region was assigned as a random variable) to determine which of seven regionalization schemes were useful in explaining region-level variance in four different population metrics.

Bass pop. metric:	Multi-level model:	Total variance	Lake-level variance	Region-level variance	ICC	F-STAT	Δ AIC	Mean region area (km^2)	Passes AIC criterion
CPUE	(EMPTY)						0.2		
	ECO 3	1.09	1.06	0.03	0.03	0.03	1.5	13996	
	ECO 4	1.08	1.08	0.00	0.00	0.00	2.2	2999	
	EDU	1.09	1.05	0.03	0.03	0.03	1.6	31636	
	HEX	1.08	1.08	0.00	0.00	0.00	2.2	649	
	HUC04	1.09	1.02	0.07	0.06	0.07	0.0	16752	
	HUC08	1.08	1.08	0.00	0.00	0.00	2.2	3046	
	HUC10	1.09	0.89	0.20	0.18	0.23	0.4	446	
Wr	(EMPTY)						0.0		
	ECO 3	70.779	64.58	6.20	0.09	0.10	0.9	13996	
	ECO 4	68.79	68.79	0.00	0.00	0.00	2.0	2999	
	EDU	68.79	68.79	0.00	0.00	0.00	2.0	31636	
	HEX	68.79	68.79	0.00	0.00	0.00	2.0	649	
	HUC04	68.796	68.75	0.05	0.00	0.00	2.0	16752	
	HUC08	68.844	67.13	1.72	0.02	0.03	2.0	3046	
	HUC10	72.47	39.52	32.95	0.45	0.83	0.4	446	
Length at Age 2	(EMPTY)						3.2		
	ECO 3	1108.3	846.80	261.50	0.24	0.31	4.5	13996	
	ECO 4	1179.4	445.80	733.60	0.62	1.65	2.0	2999	
	EDU	1020.5	850.90	169.60	0.17	0.20	4.5	31636	
	HEX	994	994.00	0.00	0.00	0.00	5.2	649	
	HUC04	1004.75	939.80	64.95	0.06	0.07	5.2	16752	
	HUC08	1048.2	454.80	593.40	0.57	1.30	2.1	3046	
	HUC10	1042.8	167.30	875.50	0.84	5.23	0.0	446	✓
Length at Age 5	(EMPTY)						4.6		
	ECO 3	1258.1	863.60	394.50	0.31	0.46	3.0	13996	
	ECO 4	886.5	886.50	330.40	0.37	0.37	5.3	2999	
	EDU	1220.3	707.40	512.90	0.42	0.73	2.0	31636	✓
	HEX	1280.2	50.60	1229.60	0.96	24.30	2.2	649	✓
	HUC04	1175.3	635.30	540.00	0.46	0.85	0.0	16752	✓
	HUC08	1238.56	1163.47	75.09	0.06	0.06	6.6	3046	
	HUC10	1236	1236.00	0.00	0.00	0.00	6.6	446	

Table 9: Each of the eight black bass population metrics evaluated in this study are listed beneath the regionalization scheme(s) that were determined to be optimal in grouping lakes for that metric when region was included as a fixed effect. An asterisk (*) indicates a population metric for which the optimality of two regionalization schemes was indistinguishable (delta AIC <2) and so the metric is listed under both schemes. LMB = largemouth bass, SMB = smallmouth bass, ECO 3 = version 3 of Omernik's ecoregion delineations, HUC04 = four-digit subregions, EDU = ecological drainage units.

EDU	HUC04	ECO3
LMB relative abundance	LMB relative weight	LMB length-at-age 2
	SMB relative abundance	LMB length-at-age 5
	SMB length-at-age 2*	SMB relative weight
	SMB length-at-age 5	SMB length-at-age 2*

Table 10: Penalized AIC scores for each species-specific population metric. Asterisks (*) denote schemes that were determined to be optimal or co-optimal under the penalized AIC criterion. Schemes highlighted in gray had the lowest penalized AIC scores (and therefore delta AIC values of 0) for the associated black bass population metric.

LMB CPUE

Scheme	AIC	No. Empty Reg.	$AIC_{penalized}$	Δ
ECO3	1463	0	1463	26
ECO4	1430	9	1448	11
EDU*	1437	0	1437	0
HEX	1433	159	1751	314
HUC04	1448	0	1448	10
HUC08	1426	14	1454	16
HUC10	1455	231	1917	480

SMB CPUE

Scheme	AIC	No. Empty Reg.	$AIC_{penalized}$	Δ
ECO3	1089	0	1089	8
ECO4	1097	15	1127	47
EDU	1083	1	1085	5
HEX	1108	177	1462	382
HUC04*	1078	1	1080	0
HUC08	1107	17	1141	61
HUC10	1100	257	1100	533

LMB Wr

Scheme	AIC	No. Empty Reg.	$AIC_{penalized}$	Δ
ECO3	1676	0	1676	7
ECO4	1699	13	1725	56
EDU	1671	0	1671	3
HEX	1704	176	2056	388
HUC04*	1668	0	1668	0
HUC08	1677	18	1713	45
HUC10	1665	258	2181	513

SMB Wr

Scheme	AIC	No. Empty Reg.	$AIC_{penalized}$	Δ
ECO3*	761	2	765	0
ECO4	778	18	814	49
EDU	773	1	775	10
HEX	769	197	1163	398
HUC04	766	2	770	5
HUC08	783	22	827	62
HUC10	741	283	1307	542

LMB Length-at-age 2

Scheme	AIC	No. Empty Reg.	$AIC_{penalized}$	Δ
ECO3*	1032	2	1036	0
ECO4	1041	28	1097	61
EDU	1043	2	1047	12
HEX	1023	217	1457	422

SMB Length-at-age 2

Scheme	AIC	No. Empty Reg.	$AIC_{penalized}$	Δ
ECO3*	736	3	742	0
ECO4	736	29	794	52
EDU	741	3	747	5.7
HEX	735	225	1185	444

HUC04	1040	1	1042	6.6
HUC08	1033	32	1097	62
HUC10	1025	307	1639	603

HUC04	738	2	742	0.1
HUC08	726	34	794	52
HUC10	727	312	1351	609

LMB Length-at-age 5

Scheme	AIC	No. Empty Reg.	$AIC_{penalized}$	Δ
ECO3*	1075	2	1079	0
ECO4	1081	32	1145	65
EDU	1077	4	1085	5.6
HEX	1055	222	1499	420
HUC04	1082	2	1086	6.7
HUC08	1078	33	1144	65
HUC10	1059	355	1769	689

SMB Length-at-age 5

Scheme	AIC	No. Empty Reg.	$AIC_{penalized}$	Δ
ECO3	671	3	677	5.8
ECO4	675	32	739	68
EDU	669	3	675	4.2
HEX	670	228	1126	455
HUC04*	667	2	671	0
HUC08	674	36	746	75
HUC10	677	318	1313	642

Table 11: Spearman rank correlations of bass population metrics averaged across the regions of three schemes that we found to be consistently optimal in explaining variance.

			Largemouth Bass				Smallmouth Bass			
			CPUE	Wr	LA2		CPUE	Wr	LA2
EDU:	**Wr**	rho	0.13			**Wr**	rho	0.15		
		n	12				n	10		
		p-value	0.68				p-value	0.68		
	LA2	rho	0.06	0.05		**LA2**	rho	-0.38	0.22	
		n	10	10			n	9	9	
		p-value	0.87	0.88			p-value	0.31	0.58	
	LA5	rho	0.12	-0.07	0.26	**LA5**	rho	-0.02	-0.07	0.08
		n	8	8	8		n	9	9	9
		p-value	0.78	0.87	0.53		p-value	0.97	0.86	0.83
HUC04	**Wr**	rho	-0.01			**Wr**	rho	-0.05		
		n	10				n	8		
		p-value	0.99				p-value	0.91		
	LA2	rho	-0.25	0.15		**LA2**	rho	-0.48	0.21	
		n	9	9			n	8	8	
		p-value	0.52	0.70			p-value	0.23	0.61	
	LA5	rho	0.45	0.05	-0.12	**LA5**	rho	-0.26	0.07	0.17
		n	8	8	8		n	8	8	8
		p-value	0.26	0.91	0.78		p-value	0.53	0.87	0.69
ECO3	**Wr**	rho	0.12			**Wr**	rho	0.25		
		n	9				n	7		
		p-value	0.77				p-value	0.59		

Largemouth Bass				
		CPUE	Wr	LA2
LA2	rho	-0.36	0.68	
	n	7	7	
	p-value	0.43	0.09	
LA5	rho	0.43	0.04	-0.14
	n	7	7	6
	p-value	0.34	0.94	0.79

Smallmouth Bass				
		CPUE	Wr	LA2
LA2	rho	-0.94	0.26	
	n	6	6	
	p-value	**0.005***	0.62	
LA5	rho	0.14	0.03	0.03
	n	6	6	6
	p-value	0.79	0.96	0.96

Figures:

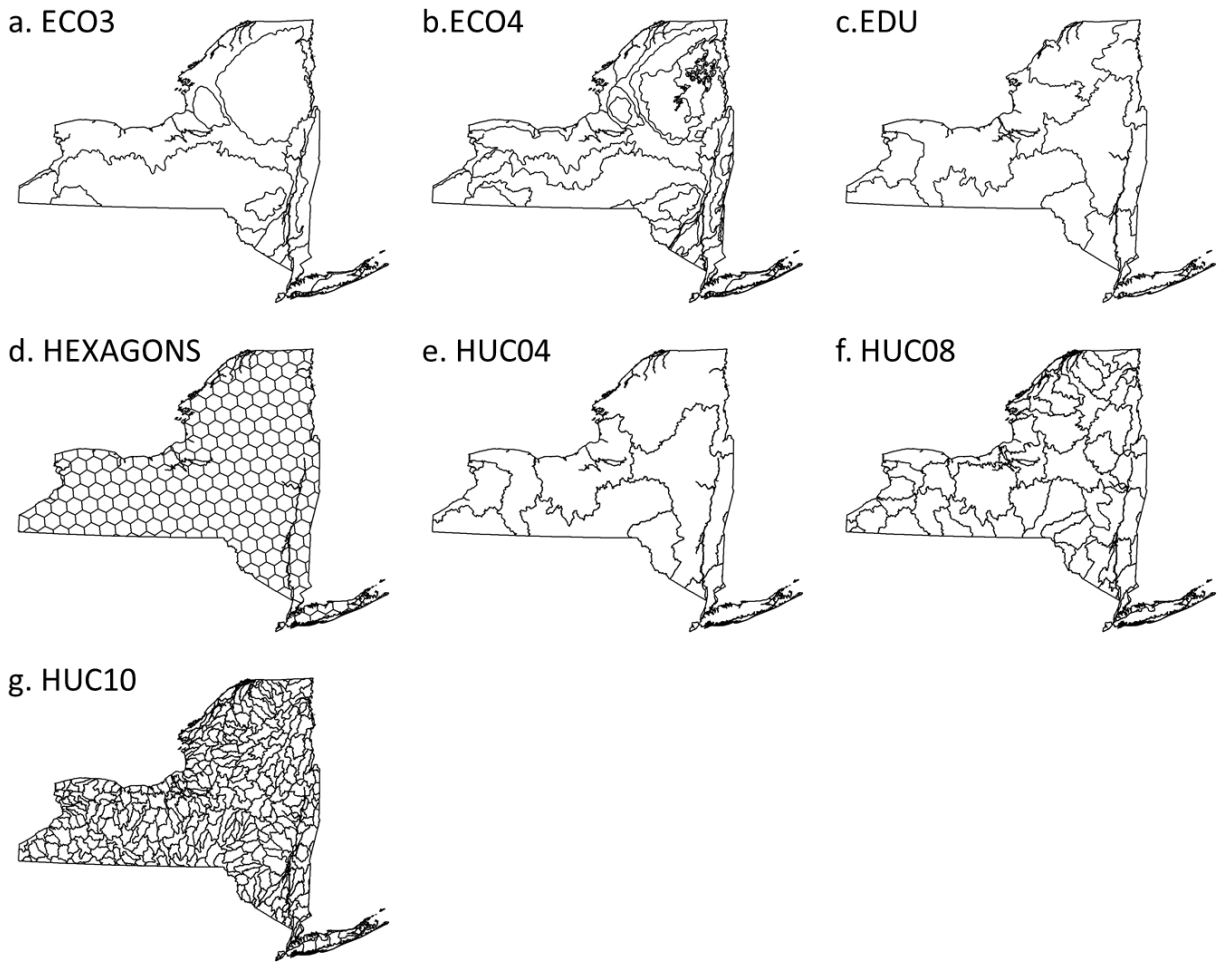

Figure 1: We evaluated the performance of 7 different regionalization schemes in explaining variability in black bass population metrics. The regionalization schemes each have relevancy to lake bass populations and included a) Omernik's ecoregion version 3, ECO3, b) Omernik's ecoregion version 4, ECO4, c) ecological drainage units, EDU, d) US EPA EMAP hexagons, HEXAGONS, e) 4-digit USGS watershed boundaries, HUC04, f) 8-digit USGS watershed boundaries, HUC08, g) 10-digit USGS watershed boundaries, HUC10. "HUC" is a widely used acronym for hydrologic unit code. Refer to Table 2 for further details about the schemes.

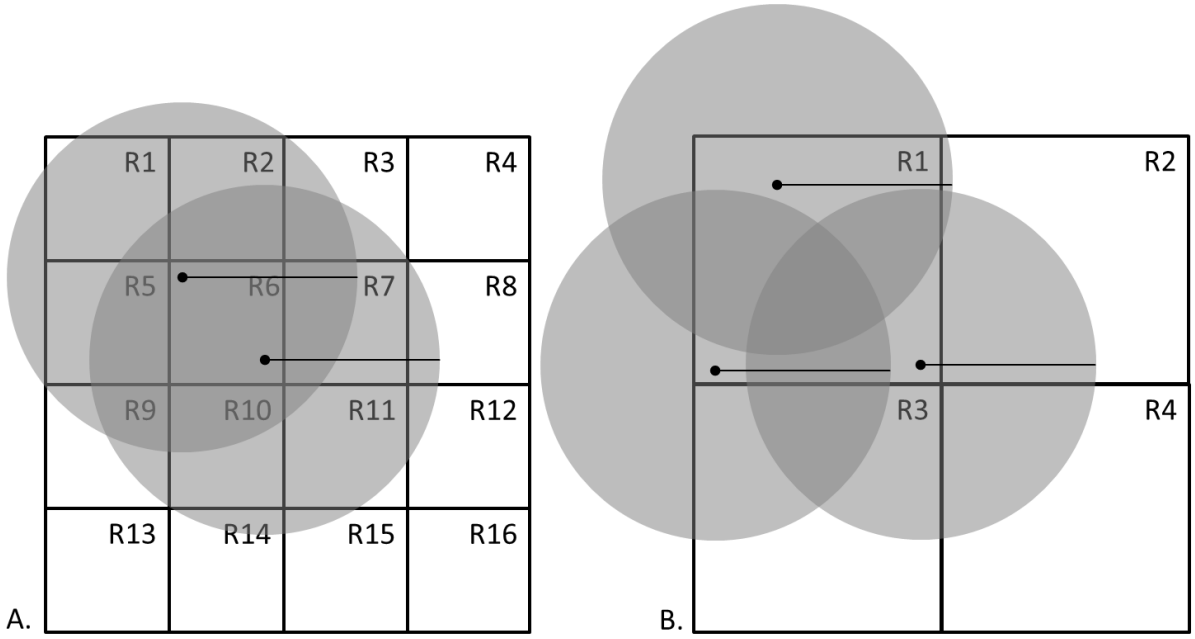

Figure 2: In 'A', regions are small enough relative to the semi-variogram range parameter (represented by the radius of the grey circles) that any two lakes (represented by black dots) within a region (represented by squares) are spatially autocorrelated resulting in low within-region variance in the response metric. In 'B', the regions are large enough relative to the same range parameter to include lakes that are not spatially autocorrelated resulting in high within-region variance in the response metric.

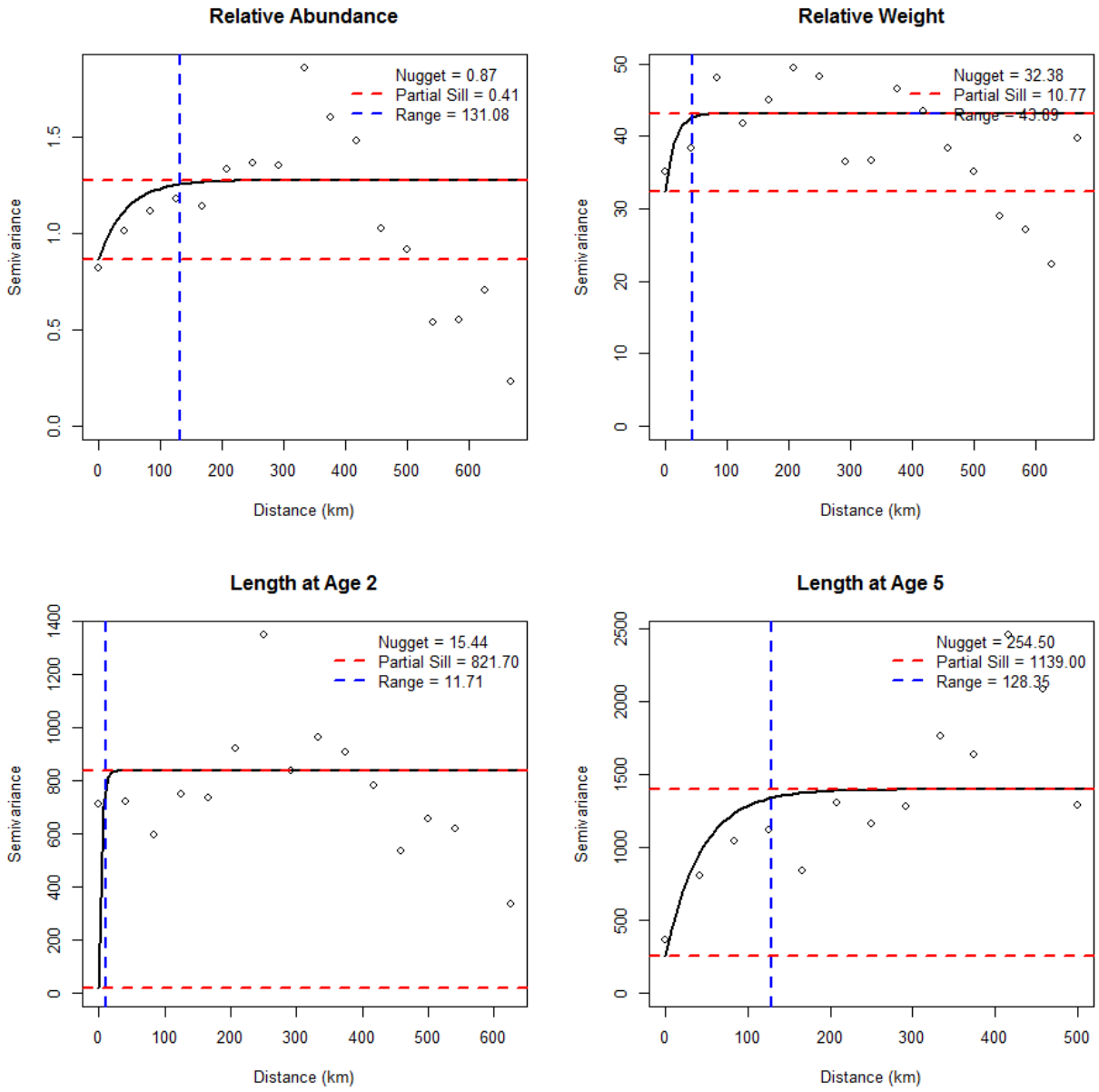

Figure 3: Semi-variogram plots and parameter estimates for four largemouth bass population metrics summarized across lakes in New York State.

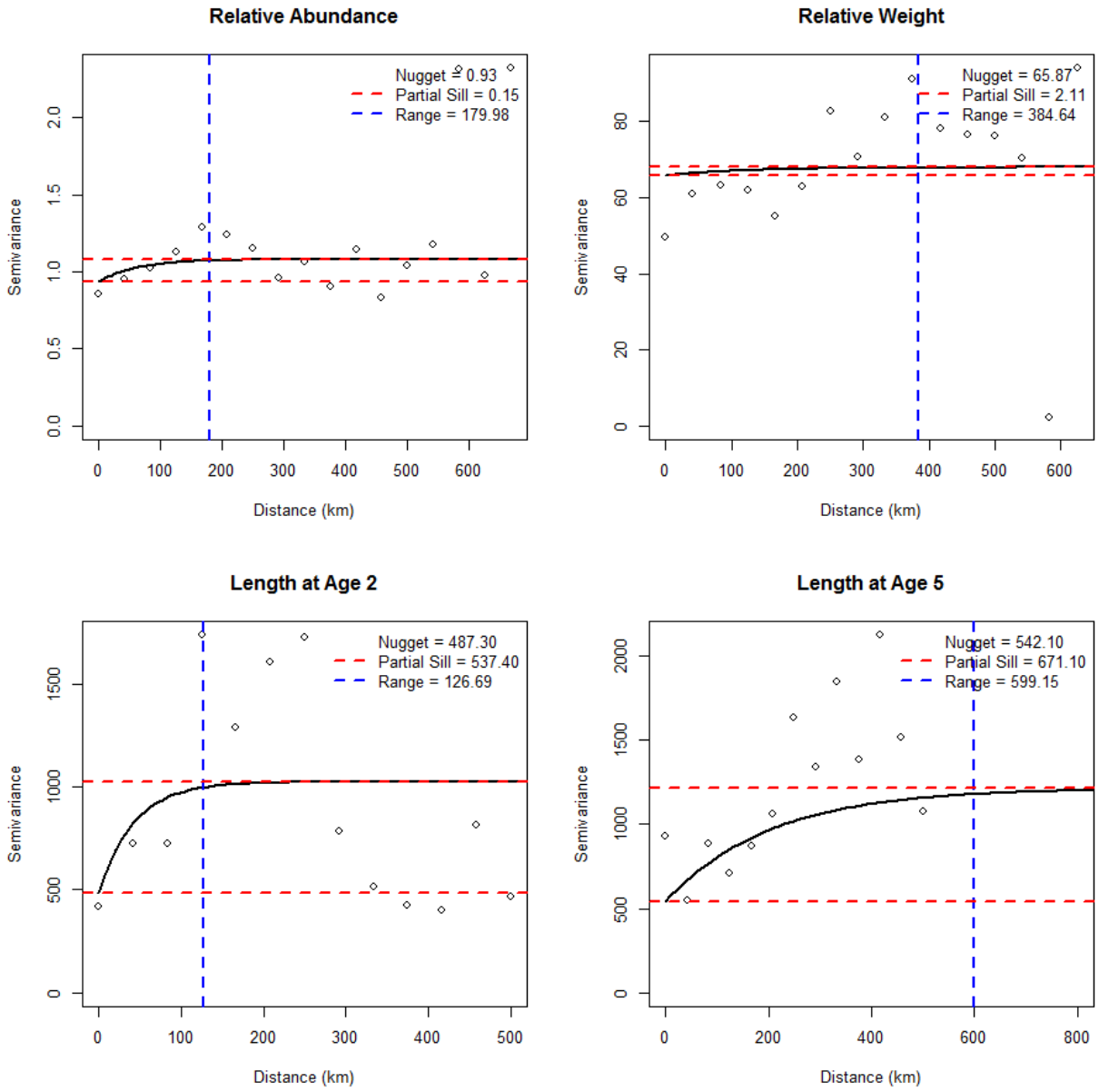

Figure 4: Semi-variogram plots and parameter estimates for four smallmouth bass population metrics summarized across lakes in New York State.

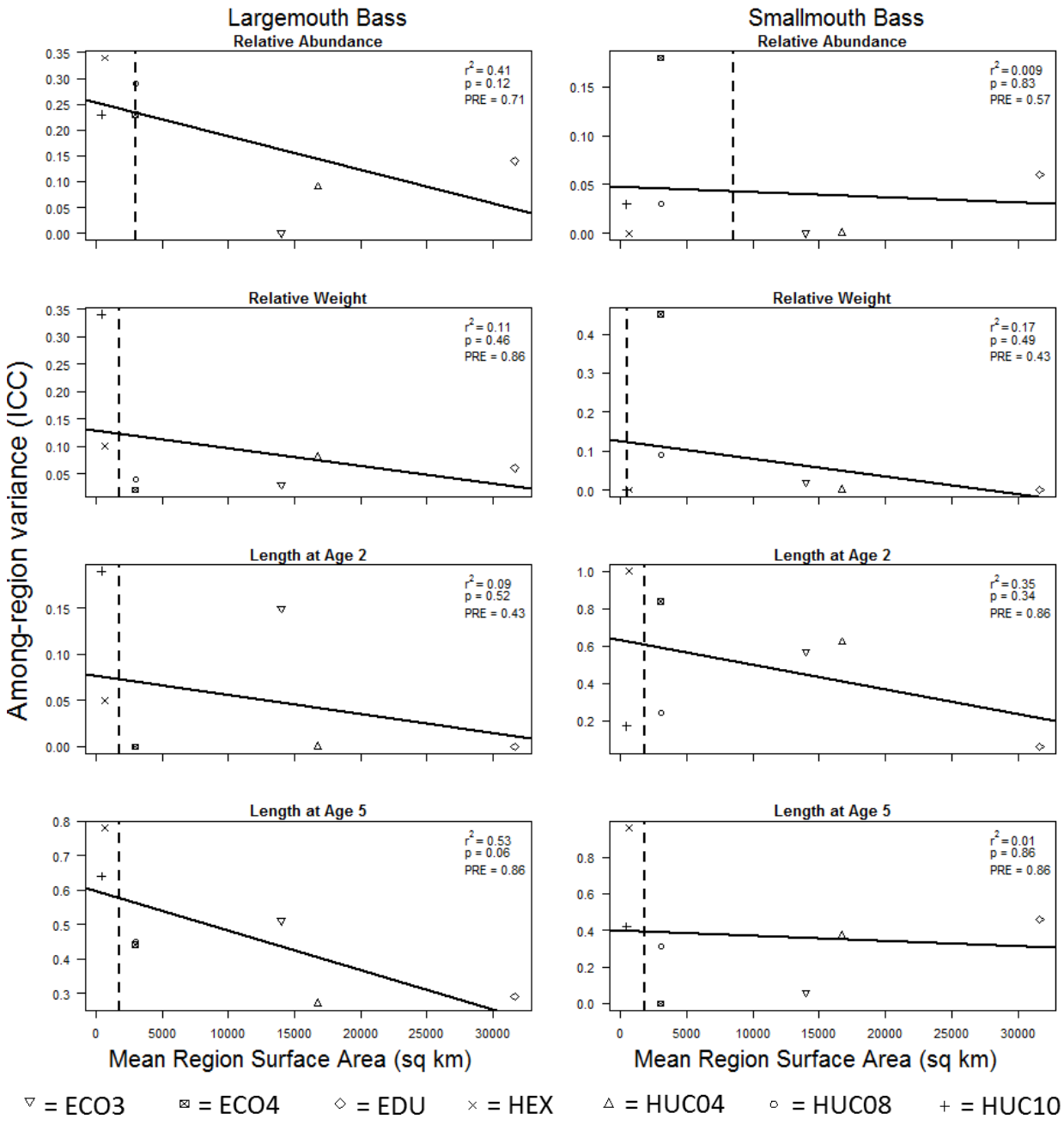

Figure 5: Regressions between among-region variance and mean region surface area for relative abundance, relative weight, length at age 2, and length at age 5 of both largemouth bass (left side) and smallmouth bass (right side). The dotted line represents the split identified by the regression tree.

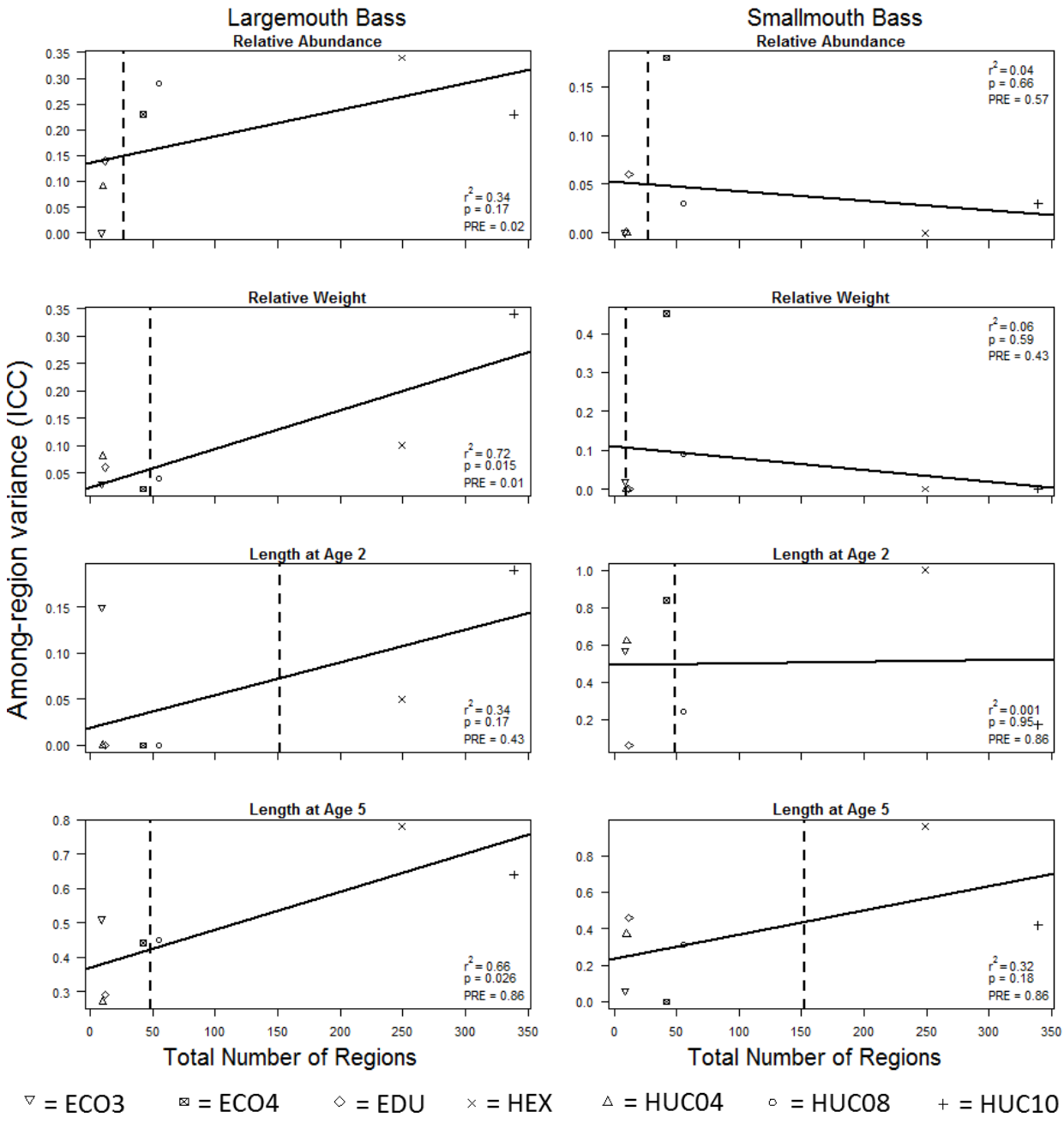

Figure 6: Regressions between among-region variance and total number of regions for relative abundance, relative weight, length at age 2, and length at age 5 of both largemouth bass (left side) and smallmouth bass (right side). The dotted line represents the split identified by the regression tree.

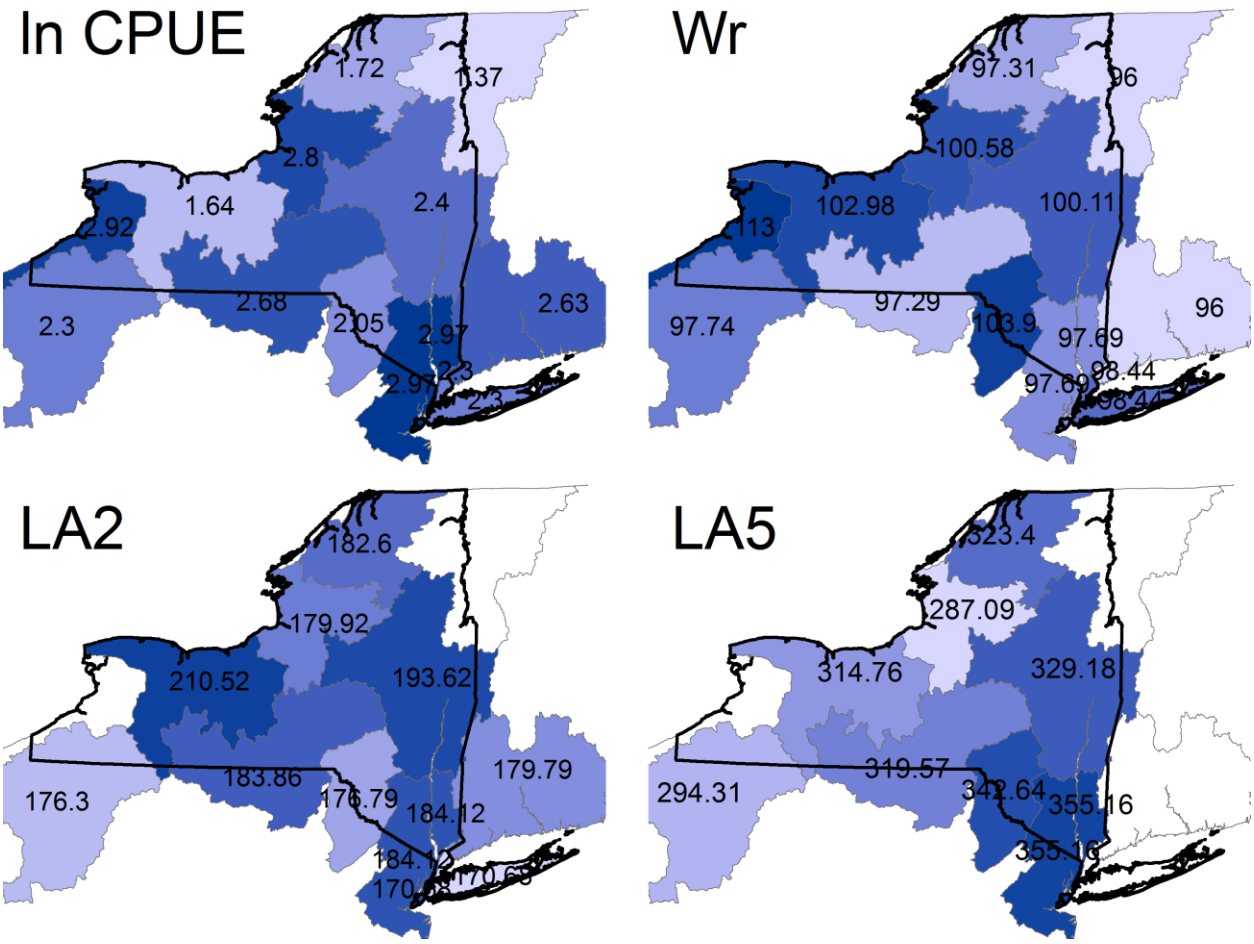

Figure 7: Maps depicting the per-region means for each of the four largemouth bass population metrics using the Environmental Drainage Unit (EDU) scheme.

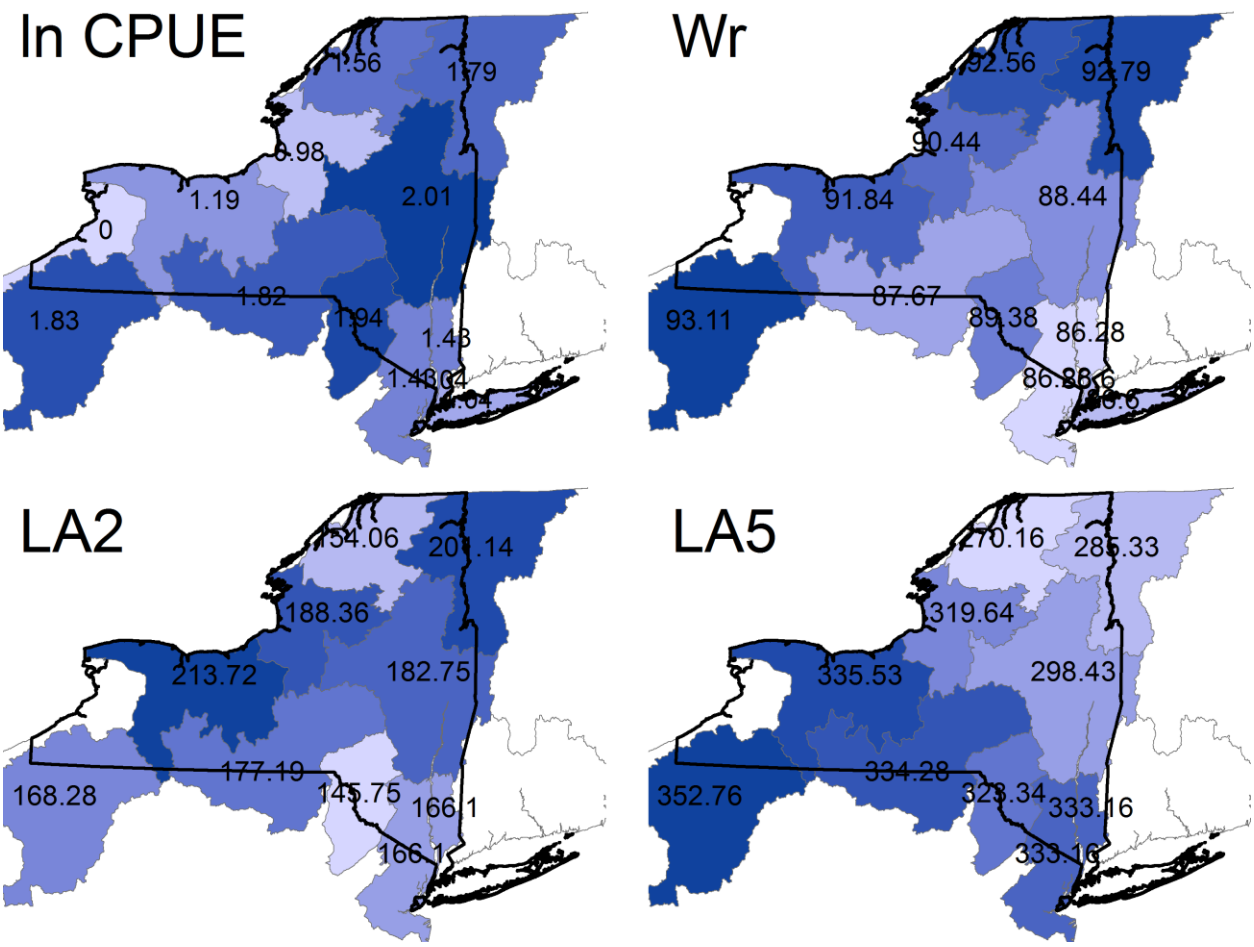

Figure 8: Maps depicting the per-region means for each of the four smallmouth bass population metrics using the Environmental Drainage Unit (EDU) scheme.

CHAPTER 2

Abstract

In chapter one, we ranked the performance of regionalization schemes in grouping black bass metrics, thus evaluating the usefulness of one scheme relative to another for bass abundance, condition, and growth metrics. In chapter two, we extend the concepts of chapter one by first identifying whether regionalization schemes group lakes for bass metrics better than random chance alone. For each bass population metric, we then evaluate the influence of relevant predictor variables including lake morphological features such as surface area, maximum depth, and elevation; catchment land cover characteristics such as proportion of the catchment that is made up of deciduous forest or cultivated fields; and climate characteristics such as mean summer temperatures and total growing degree days per year. We establish a-priori predictions of the relationship of each bass metric to each environmental characteristic. Our results show that regionalization schemes do in fact tend to group lakes better than chance alone for most bass population metrics. We also show that the coefficients of most predictors we evaluated exhibited the expected direction of the relationship (whether positive or negative). The most important predictors are easily derived from GIS systems and thus may have wide application in other contexts.

Introduction

In regions where there are hundreds or thousands of lakes, fisheries managers face a formidable challenge in developing effective and practical monitoring strategies. Typically, only a small proportion of the total number of lakes can be monitored and care must be taken to make sure that the lakes that are monitored are broadly representative of those that are unmonitored. If lakes can be grouped into strata such that the information learned from monitoring one lake

generally applies to other lakes in the same strata, managers can substantially reduce the burden of sampling many lakes while continuing to provide information useful for effective management. Grouping lakes by strata also enables managers to establish targeted fishing regulations, stocking plans, outreach programs, and educational materials that are reliably optimized and custom-tailored for the particular conditions of lakes within a given strata.

Methods for organizing lakes for fisheries management can fall under one of two broad themes. One is to group lakes using geographically contiguous regions. This method may rely on ad hoc geographic delineations such as political boundaries, or it may have ecosystem-based underpinnings such as ecoregion or watershed delineations. An alternative approach for grouping lakes is by lake-specific environmental characteristics. This second approach may rely on combinations of geographic information system (GIS)-derived data (e.g. lake surface area or land-use proportions in the catchment) and field-collected data (e.g. lake depth) describing features of each lake. Whereas the first method demands that lakes in the same strata share geographic proximity to one another (e.g. all lakes in a strata must be contained within a single continuous polygon which does not overlap that of another strata), the second method may produce strata in which member lakes are scattered geographically. The first method is the simpler of the two to implement and may be of more practical use to fisheries managers because all that is required to place a lake into a particular stratum is knowledge of its location. Once you know its geographic position, you can assign a lake to a category. In contrast, under the second approach, physical attributes of the lake or its catchment must be acquired before it can be assigned to an appropriate stratum. The strength of the first method is strongly influenced by how closely tied the features of the landscape which define the regions are to the particular lake metric of interest. In contrast, the usefulness of the second method is dependent upon the

strength of the relationships between the lake metric of interest and the specific collection of related environmental variables. Thus, there is inherently more flexibility in the second method because environmental variables thought to be important to the lake metric of interest can be hand-picked and collected. However, the second method may be more costly in terms of both time and money due to the expense associated with collecting variables.

The first approach has been used to determine how best to monitor water quality, water chemistry, and nutrient levels of lakes (Cheruvelil et al. 2008, Heiskary et al. 1987, Jenerette et al. 2002) as well as to determine macroinvertebrate assemblages (Johnson 2000), but it has not often been applied to monitoring single-species fish population characteristics (chapter one, Wagner et al. 2007) The second approach has been widely used not only to monitor fish assemblage patterns (Johnson et al. 1977, Tonn et al. 1983, Dolman 1990, Schupp 1992) but also to monitor single-species fish population demographic characteristics such as presence or absence, abundance, size structure, condition, growth and mortality (Guy and Willis 1995, Nate, et al. 2003, Paukert and Willis 2004, Tomcko and Pierce 2005, Shoup et al. 2007).

As popular sport fish, black bass regularly receive priority attention from fisheries management agencies, often being the focus of specialized state bass management plans (Jacobs et al. 1999, Jordan 2001, Hobbs et al. 2002, Bremigan et al. 2004, Anonymous 2011). In New York, there are two species of black bass – largemouth bass, Micropterus salmoides and smallmouth bass, M. dolomieu – inhabiting the majority of New York's approximately 7,800 lakes (Anonymous 2011). Recent angler surveys reveal that black bass are the most popular sport fish in New York (Connelly et al. 1997, Connelly and Brown 2009).

In chapter one, we investigated the efficacy of seven different geographically contiguous regionalization schemes in organizing lakes based upon species-specific black bass population

metrics including relative abundance, relative condition, and growth of juveniles and adults. In the present study, we not only wanted to understand the role of predictor variables in explaining variance in the bass population metrics, but in particular, our purpose was to further elicit the role of the predictors in the context of the regionalization schemes that were determined to be either simply optimal in comparison to other schemes, or quantitatively useful (having passed multiple criteria established to gauge how well the schemes organized the bass metrics) in clustering lakes for bass populations.

Methods

To accomplish this objective, we incorporated eleven different predictor variables into multi-level models where regionalization scheme was either fixed or random (Table 12). Our models were either 2-level (when region was included as a fixed-effect) or 3-level (when region was included as a random-effect). As such, we were able to evaluate the influence of predictor variables at the survey (level 1), lake (level 2), or region (level 3) levels.

Predictions

Prior to model construction, we developed level-specific a priori expectations of the sign of the relationship between each bass metric and each predictor variable (Table 13). By and large, our predictions were strongly influenced by the belief that the primary drivers of bass population metrics are temperature and nutrient loadings and this belief is reflected in the covariate-specific predictions we describe.

At the survey level, we predicted that the year of the survey would be positively associated with the cpue and growth metrics of both species of bass. The general warming climate trend over the 24 years of our dataset (Walther et al. 2002) should be favorable to bass abundance and growth (Zweifel et al 1999), particularly in New York which is situated at the northern extent of the range of both species. We predicted that mean summer temperatures and

degree days, each averaged over a bass's lifespan, would each be positively associated with the growth metrics of both species. We also predicted that the mean summer temperatures and degree days associated with the birth years for both species of bass would be positively associated with the cpue metric. Moreover, we predicted that these climate variables during the year of capture would positively associated with the relative weight metrics of both species.

At the lake and region levels, we reasoned that the relationship between bass metrics and maximum lake depth could be either positive or negative. Although deep lakes will have colder regions which are not favorable to bass, maximum lake depth alone in not necessarily related to how much littoral habitat is available. Similarly, we predicted that lake surface area could be positively or negatively related to bass metrics because lakes with large surface areas could be either deep or shallow which would influence the temperature of the water as well as lake productivity. We predicted that bass metrics would be positively associated with lake SDI because lakes with more shoreline would tend to provide more littoral habitat for both species of bass. We predicted a negative relationship between largemouth bass metrics and lake elevation due to the fact that higher lakes would have relatively cooler water. Since smallmouth bass are more tolerant of cooler water, we predicted that lake elevation could be either positively or negatively associated with smallmouth bass metrics. We used similar reasoning to anticipate negative relationships between largemouth bass metrics and lake latitude, but either positive or negative relationships between lake latitude and smallmouth bass metrics. We included longitude as a predictor variable because of its potential power in detecting broad east-west trends, but we did have any reason to predict either positive or negative relationships between longitude and bass metrics. We reasoned that increasing proportions of deciduous forest in the catchment area surrounding a lake would be associated with lower nutrient loading and thus

negative relationships with bass abundance and growth, but positive relationships could result, particularly for smallmouth bass, since deciduous forest could imply a more balanced ecosystem and might also result in cooler water entering the lake. In contrast, we reasoned that increasing proportions of cultivated crops in lake catchment areas would be associated with higher nutrient loadings and warmer waters, thus we predicted positive relationships between this predictor and largemouth bass metrics, but either positive or negative relationships with smallmouth bass metrics. As with the survey level, we predicted that the mean summer temperatures and degree days associated with surveys averaged across lakes would both be positively associated with the abundance and growth metrics of both species of bass.

Although we suspected that the influence of landscape level factors on bass condition might be less than on bass abundance or growth, we expected the relationships between the predictors and bass relative condition to generally mirror the relationships with bass growth. Moreover, our predictions as to the sign of the relationships between bass metrics and environmental predictor variables were agnostic regarding the influence of biotic interactions such as predation and competition.

Fisheries Surveys and Population Metrics

Fisheries surveys of lakes across the state of New York were conducted between 1988 and 2011 according to standardized protocols (Green 1989). From the data collected during these surveys, the following four population metrics were calculated for both largemouth bass and smallmouth bass: relative abundance (the number of bass collected in one hour of night electro-fishing), relative condition (relative weight), juvenile growth (length at age 2), and adult growth (length at age 5). The four population metrics were calculated independently for each species of bass resulting in a total of 8 metrics. A full description of survey methods including gear types used, minimum sample sizes, and criteria for data inclusion as well as a description of

metric derivations are provided in chapter one.

Environmental Variables

Environmental variables describing lake morphometric features (surface area, shoreline development index (SDI), maximum depth, latitude, longitude, and elevation), catchment land cover characteristics (percent cultivated crops and percent deciduous forest), and climate (mean annual air temperature and degree days) were summarized for each lake (Table 12).

All lake morphometric variables other than maximum depth were derived from layers in a geographic information system. SDI, which relates the measured shoreline length of a given lake to the shoreline length of a perfectly circular lake of equal area was computed using (Equation 1). The source of the vector data (lake polygons) in the geographic information system was the national hydrography dataset, NHD (Simley and Carswell 2009). The raster data from which lake elevations were derived came from the U.S. Geological Survey's GAP analysis program (USGS GAP 2011). The lake depth data came from either regional fishing map guides or lake contour maps available on the New York Department of Environmental Conservation website.

$$SDI = \frac{perimeter}{2\sqrt{surface\ area * \pi}} \qquad \textbf{Equation 1}$$

Geographic information system tools were used to delineate catchments of all tributaries delivering water to each lake and to determine the proportion of land cover types for each lake's tributary catchment. Headwater lakes which, by definition have no tributary inputs, also have no tributary catchments and thus we set the proportion of all land cover types for headwater lakes to zero. The source of the land cover data was the national land cover dataset (NLCD 2006) (Fry et al. 2011).

Monthly surface air temperature data was downloaded from the National Climatic Data Center (NCDC) of the National Oceanic and Atmospheric Administration (NOAA) in the form of climate station point data (n = 218 stations). Annual summer (June through August) temperature means were computed for each climate station and geographic information system tools were used to interpolate climate station data across the state of New York and assign mean summer annual air temperature values to the center point of each lake. Annual degree days (sum of the number of degrees above 10°C each day for the entire year) were computed for each lake from spatial interpolations of the same climate station data. Recognizing that growth at a given age is additive across all growing seasons of a bass's lifetime, for each bass growth record, we averaged the mean summer air temperatures across all years of the bass's lifetime and we also averaged the annual degree days over the bass's lifetime.

Finally, the year in which the survey was conducted was also included as a numerical fixed-effect explanatory variable in order to determine if the bass population metric of interest was increasing or declining linearly over time across all lakes.

Statistical Analysis

In chapter one, we identified from our models where region was included as a fixed effect which of seven regionalization schemes optimally grouped lakes according to several key black bass population metrics, but it remained unclear whether organizing lakes according to those schemes was a useful way to explain variance in black bass population metrics, especially in comparison to the use of traditional environmental covariates such as lake depth, surface area, and catchment characteristics. Here, we sought to further elucidate the usefulness of regionalization schemes in comparison to traditional environmental covariates in explaining variance in black bass population metrics. To do so, for each population metric, we included the regionalization scheme that we had earlier identified as optimal as a fixed covariate in new 2-

level models which also included the additional predictors. We used a bottom-up approach to run the 2-level models. To be clear about the thought process that went into the inclusion of predictors, we include a table of a priori expectations of the relationships between the predictors and the responses (Table 13).

Two-level model-building approach:

In preparation for the 2-level bottom-up model-building process, we first re-ran the model with only regionalization scheme as a (fixed effect) predictor variable and computed complete pairwise Tukey comparisons of all regions within the scheme. If any pairs of regions exhibited significant differences, we marked that scheme for inclusion in future models. In contrast, if there were no significant differences among all pairwise region comparisons, we considered the scheme to be ineffective in explaining variance in the population metric despite its earlier selection as the optimal scheme relative to all other schemes.

For metrics where the regionalization scheme was considered to be ineffective, we still wanted to know how well environmental variables performed in explaining variance. Therefore, we ran separate models which included each environmental predictor, one at a time. Predictor variables were cluster-centered according to the level of the model to which they were relevant. This standardization allowed for easier interpretation of results. Following the general model-building principles set forth in Gelman and Hill (2007), if an environmental predictor was significant and had the expected sign, we marked it for inclusion in future models. If it was significant, but did not have the expected sign, we thought carefully about what might be driving the unexpected sign and if the inclusion of other predictor variables or interactions might further inform that unexpected relationship. Similarly, if the predictor was not significant but had the expected sign, we marked it for inclusion in future models, but if the predictor was not significant and did not have the expected sign, we removed it from further consideration. We

then ranked all the retained predictors based on their perceived importance in relation to the population metric. Then we built a new model which included the most important predictor along with the second-most important predictor. If the new model had an AIC score that was less than the model with the most important predictor alone, we retained both predictors in further models; otherwise, we discarded the second-most important predictor and tried again with the third-ranked predictor. We continued in this stepwise fashion until we had evaluated all remaining predictors. Finally, we systematically evaluated interactions. First, we made a list of potential interactions. Then, we added each one individually to the model to determine which interactions lowered the AIC score. We then ranked the interactions which lowered the AIC score and followed a stepwise procedure of adding the top-ranked interaction to the base model, evaluating whether it further lowered the AIC score, then adding in the next interaction and evaluating once more. Interactions that did not lower the AIC score were thrown out.

In contrast, for metrics where the regionalization scheme had some pairs of regions that were significantly different, we followed the same approach to continued model-building as described above except that for these models, scheme was included throughout the model-building process.

For models where region was included as a fixed effect or was not included at all, the structure of the model was necessarily two-level, the first level represented by survey and the second level represented by lake. Since region was considered to be a fixed effect, it could not be included as a third level in these models.

Three-level model-building approach:

For models where region was included as a random effect, we followed a bottom-up model-building strategy for the inclusion of predictor variables. The structure of these models was necessarily three levels, the first level being survey, the second level lake, and the third level

region. Instead of building 3-level models for all eight black bass metrics, we focused on only the two bass metrics for which our earlier study had determined that a particular regionalization scheme was convincingly effective in partitioning the variance. Thus, the two bass metrics we built 3-level models for were largemouth bass relative abundance and largemouth bass length at age five.

When we initially evaluated a predictor variable in a model by itself, we considered its inclusion at all levels of the model again using the Gelman and Hill (2007) guidelines to determine which levels of any given predictor to include in subsequent models. For example, in a three level model where the response variable was largemouth bass relative abundance, we considered that maximum lake depth was an environmental variable that could vary at the lake level and at the region level, but not at the survey level. Therefore, in our evaluation of maximum lake depth, we first ran a model where maximum lake depth was included at the lake level (using region-centered observations). We then ran a second model where maximum lake depth was included only at the region level (using grand mean-centered observations). And finally, we ran a third model where maximum lake depth was included at both the lake and region level. We then used Gelman and Hill's (2007) guidelines to determine which levels of maximum lake depth to retain for future models.

Results

Two-Level Bottom-Up Models:

In the 2-level models where region was included as a fixed effect categorical predictor variable, and model-building was from the bottom-up, we found that for six of the eight bass population metrics evaluated, regions grouped lakes better than would be expected from random

chance alone (Table 14). For the five bass metrics where region was significant, the inclusion of additional predictors was generally useful in explaining variance in the response, the exception being smallmouth bass relative weight where none of the predictor variables we evaluated were found to be useful (Table 15). The inclusion of additional predictors was also useful in explaining variance for the two bass metrics where region did not group lakes better than would be expected from random chance alone (largemouth bass length at age 2, and smallmouth bass length at age 5) (Table 16, Table 17). For the five bass metrics which benefitted from the addition of both region and other predictors, the inclusion of predictor variables reduced the total variance between 11% (smallmouth bass cpue) and 33% (smallmouth bass length at age 2), the mean (+/-SD) reduction being 21% (+/- 8%) (Table 16, Table 17).

The predictors that were most often included in final models were elevation and proportion of catchment in cultivated crops (used in 4 metrics each) and summer temperature, maximum depth, and surface area, (used in 3 metrics each) (Table 15). The signs of the coefficients for these predictors were generally in agreement with our a priori hypotheses (Table 13, Table 15). For example, the coefficients for elevation were negative for all models where that predictor was included (Table 15, Table 16, Table 17). And the coefficient for proportion of catchment in cultivated crops was positive in three of the four models it was included in (Table 15, Table 16, Table 17). Across the seven models which included predictor variables, 23 of the 26 predictor coefficients (eighty-eight percent) had the expected sign (Table 15).

Three-Level Bottom-Up Models:

Predictor variables from all three levels were found to be significant in the final 3-level models for both largemouth bass CPUE and largemouth bass length at age 5 (Table 18, Table 19). For the largemouth bass length at age 5 model, the signs of the relationships between

response and each predictor were always as expected (Table 18, Table 19). However, for the model where largemouth bass CPUE was the response, the signs of the coefficients for half of the response-predictor relationships were unexpected. In particular, the sign of the survey-level predictor, degree day, was unexpectedly negative and the sign of the SDI predictor was unexpectedly negative at both the lake and region levels of the model (Table 18, Table 19). Survey year and lake elevation were significant predictors in the final 3-level models for both bass metrics and the sign of the relationship for each of those predictors was as expected in all cases (Table 18). The addition of predictor variables reduced the total variance by 15% in the case of largemouth bass cpue and a substantial 47% in the case of largemouth bass length at age 5 (Table 18).

Discussion

We answered two fundamental questions in this analysis: 1) Do the regions of the regionalization scheme group the bass metric better than would be expected from random chance alone? 2) Which predictor variables are helpful in explaining remaining variance in the bass population metric?

Is regionalization scheme a useful way to group lakes for bass population metrics?

The results of our 2-level bottom-up models suggest that the regions of the regionalization scheme grouped bass metrics better than would be expected from random chance alone for seven out of the eight bass metrics. For each of the eight bass metrics, the regionalization scheme that we evaluated was the scheme that we had identified in chapter one to be the optimal of seven different schemes for that specific metric. Despite the evidence that regionalization schemes group lakes for bass population metrics better than chance alone, it's important to recognize that even in cases where quantitative criteria are satisfied, a scheme still

may not cluster lakes in a way consistent with biological or ecological differences.

The utility of the specific regionalization schemes we used in our models for grouping lakes for bass metrics is tied to the way the schemes delineate broad landscapes based upon ecosystem and hydrologic characteristics that are important to bass. For example, the ecological drainage unit (EDU) scheme which we found to be helpful in grouping lakes for largemouth bass CPUE was delineated to conform to patterns in physiography, climate, and freshwater ecosystem connectivity (Higgins et al. 2005). The ecoregion (Omernik, version 3) scheme which we found to be helpful in grouping lakes for largemouth bass length at age 5, smallmouth bass relative weight and smallmouth bass length at age 2, was delineated based on observed patterns of land use, land surface form, natural vegetation, and soils (Omernik 1987). And the four-digit hydrologic unit code (HUC04) scheme that we found helpful in grouping lakes for largemouth bass relative weight and smallmouth bass length at age 5, divided the landscape based upon surface water drainage areas (Seaber et al. 1987).

The role of predictor variables

Temperature and nutrient loadings are known to be important drivers of bass abundance and growth. Therefore, we predicted that the three survey-level predictors we evaluated (survey year, mean summer temperature, and degree days) would be positively related to each of the eight bass metrics we investigated. In our 2-level bottom-up models, we found that survey year was indeed positively related to smallmouth bass CPUE, but was negatively related to largemouth bass relative weight. We also found that mean summer temperature was positively related as expected to length at age 5 for both species as well as length at age 2 for smallmouth bass. In our 3-level bottom-up models, we found that survey year was positively related to the two bass metrics we evaluated, but degree days were unexpectedly negatively related to largemouth bass CPUE.

The only other significant response-predictor relationships in the 2-level bottom-up and 3-level bottom up models that was unexpected was the significant negative relationship between largemouth bass cpue and shoreline development index. Wagner et al. (2007) showed a similar relationship between yellow perch length at age 2 and shoreline development factor. While cpue and catchability are not strictly speaking the same thing, studies on catchability may shed light on this unexpected negative relationship. For example, Schoenebeck and Hansen (2005) showed that spring electrofishing catchability of largemouth bass in lakes in Wisconsin was negatively related to SDI (coefficient = -1.843, p-value = 0.0004), but there was no significant relationship between fall electrofishing catchability of largemouth bass and SDI (coefficient = -0.387, p-value = 0.314). In contrast, neither spring nor fall electrofishing catchability was significantly related to SDI for smallmouth bass in Wisconsin lakes (spring coefficient = -1.181, p-value = 0.064; fall coefficient = -0.837, p-value = 0.856). Largemouth bass are dependent upon shallow littoral habitats. In the spring and early summer, largemouth bass migrate toward shore to build and guard nests for reproduction. If a lake's shoreline is complex, springtime electrofishing survey boats may not be able to access important near-shore concentrations of largemouth bass. With their lower thermal tolerances, smallmouth bass concentrations tend to be lower in the water column, further away from the shore than largemouth bass and, thus, springtime electrofishing surveys would not be as restricted in their access to smallmouth bass concentrations. Schoenebeck and Hansen (2005) suggest a biological basis for the observed negative relationship between largemouth bass relative abundance and lake SDI might be that lakes with lower SDI have steeper banks, "thereby concentrating fish in the available littoral habitat during spring spawning, which results in higher spring catchability estimates".

In cases such as ours where there are a large number of predictors, there is a lack of

consensus in the literature as to whether to build regression models from the bottom-up (adding and evaluating one predictor at a time) as Wagner et al. (2007) did, or from the top-down (including all predictors and potential interactions in a single "full" model and then eliminating predictors until the best model is achieved) as Zuur et al. (2009) and Grueber et al. (2011) recommend. Advocates of the bottom-up approach point out that it is more clearly hypothesis-driven. In the bottom-up approach, predictor variables are only added to the model if there is a clear expectation as to why they might have an influence on the response variable. In contrast, the top-down approach has been criticized as having the potential to be a "thoughtless approach". Burnham and Anderson (2002) caution that "Letting the computer find out is a poor strategy and usually reflects the fact that the researcher did not bother to think clearly about the problem of interest and its scientific setting".

Believing the arguments for the bottom-up approach to be more convincing, we opted to run the 2-level models using the bottom-up method. The results that we present in this paper reflect that choice, but for the sake of comparison, in a parallel analysis (see supplementary materials), we also ran the models using the top-down approach. In doing so, we determined that the bottom-up model-building strategy described by Wagner et al. (2007) produced results that were similar, albeit, more concise, concentrated and clear that the results produced by the top-down model-building strategy suggested by Zuur et al. (2009) and Grueber et al. (2011). While both strategies identified roughly the same significant predictor variables for all eight bass models, the top-down strategy usually retained several additional predictors that were not significant. The bottom-up strategy identified the predictors that were most strongly associated with the response. For this reason, we suggest the bottom-up model-building strategy be used for future similar analyses.

Our results suggest that three key predictor variables to include in models designed to explain variation in lacustrine bass abundance, growth or condition metrics are lake surface area, lake elevation, and the proportion of lake catchment that is covered in cultivated crops. Fortunately, all of these predictors are relatively simple to acquire using geographic information systems tools so we recommend that these predictors be included whenever possible in models attempting to explain variance in these bass metrics.

Despite the fact that their results failed to show that ecoregions and watersheds were effective in grouping lakes for warmwater fish growth rates, Wagner et al. (2007) used methodology similar to ours to demonstrate that local lake characteristics including lake surface area, mean lake depth, shoreline development factor, and growing degree days can explain a significant amount of variation in the mean length at age of warmwater fish species. Our results offer further support of those findings.

References

Anonymous (2011). Florida Black Bass Management Plan. Florida Fish and Wildlife Conservation Commission. 53 pp. Retrieved from (http://myfwc.com/fishing/freshwater/black-bass/).

Beamsderfer, R.C.P., and J.A. North (1995) Growth, natural mortality, and predicted response to fishing for largemouth bass and smallmouth bass populations in North America. North American Journal of Fisheries Management, 15, 688-704.

Bremigan, M., J. Breck, N. Godby, S. Hanshue, R. Moody, T. Rozich, and M. Thomas (2004).

Black bass fishing seasons in Michigan: background, research review, and recommendations. Michigan Department of Natural Resources Fisheries Division Report. 1-53.

Cheruvelil, K.S., P.A. Soranno, M.T. Bremigan, T. Wagner, S.L. Martin (2008). Grouping lakes for water quality assessment and monitoring: the roles of regionalization and spatial scale. Environmental Management, 41, 425-440.

Connelly, N.A., T.L. Brown, and B.A. Knuth (1997). New York Statewide Angler Survey 1996, Report 1: Angler effort and expenditures. New York State Department of Environmental Conservation, Bureau of Fisheries, Albany, NY.

Connelly N.A., and T.L. Brown. (2009). New York Statewide Angler Survey, 2007. New York State Department of Environmental Conservation, Bureau of Fisheries, Albany, NY.

Dolman, W.B. (1990). Classification of Texas reservoirs in relation to limnology and fish community associations. Transactions of the American Fisheries Society, 119, 511-520.

Fry, J., G. Xian, S. Jin, J. Dewitz, C. Homer, L. Yang, C. Barnes, N. Herold, and J. Wickham (2011). Completion of the 2006 National Land Cover Database for the Conterminous United States, Photogrammetric Engineering and Remote Sensing, 77, 858-864.

Gelman, Andrew and Jennifer Hill. 2007. Data analysis using regression and

multilevel/hierarchical models. Cambridge University Press. pg 69.

Gelman, A. and Y. Su. (2014). arm: Data Analysis Using Regression and Multilevel/Hierarchical Models. R package version 1.6-10. http://CRAN.R-project.org/package=arm.

Green, D.M. (1989). Centrarchid Sampling Manual. New York State Department of Environmental Conservation, Bureau of Fisheries, Albany, NY.

Guy, C.S., and D.W. Willis (1995). Population characteristics of black crappies in South Dakota waters: a case for ecosystem-specific management. North American Journal of Fisheries Management. 15, 754-765.

Hayes, D.B., C.P. Ferreri, and W.W. Taylor (1996). Linking fish habitat to their population dynamics. Canadian Journal of Fisheries and Aquatic Science, 53, 383-390.

Heiskary S.A., C.B. Wilson, D.P. Larsen (1987). Analysis of regional patterns in lake water quality: using ecoregions for lake management in Minnesota. Lake and Reservoir Management, 3, 337–344.

Higgins, J.V., M.T. Bryer, M.L. Khoury, and T.W. Fitzhugh (2005). A freshwater classification approach for biodiversity conservation planning. Conservation Biology, 19, 432-445.

Hinch, S.G., N.C. Collins, and H.H. Harvey (1991). Relative abundance of littoral zone fishes: biotic interactions, abiotic factors, and postglacial colonization. Ecology, 72, 1314-1324.

Hobbs, B., C. Horton, L. Claybrook, B. Shinn, D. Swann (2002). Largemouth bass management plan. Arkansas Game and Fish Commission Report, 1-36.

Idrisi, N., Mills, E. L., Rudstam, L. G., & Stewart, D. J. (2001). Impact of zebra mussels (Dreissena polymorpha) on the pelagic lower trophic levels of Oneida Lake, New York. Canadian Journal of Fisheries and Aquatic Sciences, 58(7), 1430-1441.

Jacobs, R.P., W.A. Hyatt, and E.B. O'Donnell (1999). A management plan for bass in Connecticut waters and recommendations for other warmwater species. State of Connecticut Department of Environmental Protection. Federal Aid in Sport Fish Restoration F-57-R: Job 15 Final Report, 1-47.

Jenerette G.D., J. Lee, D.W. Waller, R.E. Carlson (2002). Multivariate analysis of the ecoregion delineation for aquatic systems. Environmental Management, 29, 67–75.

Johnson R.K. (2000). Spatial congruence between ecoregions and littoral macroinvertebrate assemblages. Journal of the North American Benthological Society, 19, 75–486.

Johnson, M.G., J.H. Leach, C.K. Minns, and C.H. Olver (1977). Limnological characteristics of Ontario lakes in relation to associations of walleye (Stizostedion vitreum vitreum),

northern pike (Esoc lucius), lake trout (Salvelinus naymaycush), and smallmouth bass (Micropterus dolomieui). Journal of the Fisheries Board of Canada, 34, 1592-1601.

Jordan, R.M. (2001). Black bass management plan. Maine Department of inland fisheries and wildlife, division of fisheries and hatcheries Report, 1-39.

Krakowiak, P. J., & Pennuto, C. M. (2008). Fish and macroinvertebrate communities in tributary streams of eastern Lake Erie with and without round gobies (Neogobius melanostomus, Pallas 1814). Journal of Great Lakes Research, 34(4), 675-689.

Ludsin, S.A., and D.R. DeVries (1997). First-year recruitment of largemouth bass: the interdependency of early life stages. Ecological Applications, 7, 1024-1038.

Miranda, L.E., and L.L. Pugh (1997). Relationship between vegetation coverage and abundance, size and diet of juvenile largemouth bass during winter. North American Journal of Fisheries Management, 17, 601-610.

Nate, N.A., M.A. Bozek , M.J. Hansen , C.W. Ramm, M.T. Bremigan and S.W. Hewett (2003). Predicting the occurrence and success of walleye populations from physical and biological features of northern Wisconsin lakes. North American Journal of Fisheries Management. 23, 1207-1214.

Paukert, C.P., and D.W. Willis. (2004). Environmental influences on largemouth bass

Micropterus salmoides populations in shallow Nebraska lakes. Fisheries Management and Ecology. 11, 345-352.

Rice, J.A., J.E. Breck, S.M. Bartell, and J.F. Kitchell (1983). Evaluating the constraints of temperature, activity and consumption on growth of largemouth bass. Environmental Biology of Fishes. 9, 263-275.

Rose, K.A., J.H Cowan Jr., K.O. Winemiller, R.A. Myers, and R. Hillborn. Compensatory density dependence in fish populations: importance, controversy, understanding and prognosis. Fish and Fisheries, 2, 293-327.

Schoenebeck, C.W. and M.J. Hansen (2005). Electrofishing catchability of walleye, largemouth bass, smallmouth bass, northern pike, and muskellunge in Wisconsin lakes. North American Journal of Fisheries Management, 25, 1341-1352.

Schupp, D.H. (1992). An ecological classification of Minnesota lakes with associated fish communities. Minnesota Department of Natural Resources Investigational Report, 417, 1-27.

Shoup, D.E., S.P. Callahan, D.H. Wahl, and C.L. Pierce (2007). Size-specific growth of bluegill, largemouth bass and channel catfish in relation to prey availability and limnological variables. Journal of Fish Biology. 70, 21-34.

Simley, J.D., W.J. Carswell Jr., (2009). The National Map—Hydrography: U.S. Geological

Survey Fact Sheet 2009-3054. Retrieved from (http://pubs.usgs.gov/fs/2009/3054/pdf/FS2009-3054.pdf).

Steinhart, G.B., R.A. Stein, and E.A. Marschall (2004). High growth rate of young-of-the-year smallmouth bass in Lake Erie: a result of the round goby invasion? Journal of Great Lakes Research. 30, 381-389.

Tomcko, C.M., and R.B. Pierce (2005). Bluegill recruitment, growth, population size structure, and associated factors in Minnesota lakes. North American Journal of Fisheries Management. 25, 171-179.

Tonn, W.M., J.J. Magnuson, A.M. Forbes (1983). Community analysis in fishery management: an application with northern Wisconsin lakes. Transactions of the American Fisheries Society, 112, 368-377.

US Geological Survey, Gap Analysis Program (GAP). (2011). National Land Cover, Version 2. Retrieved from (http://gapanalysis.usgs.gov/gaplandcover/data/).

Wagner T., M.T. Bremigan, K.S. Cheruvelil, P.A. Soranno, N.A. Nate, J.E. Breck (2007). A multilevel modeling approach to assessing regional and local landscape features for lake classification and assessment of fish growth rates. Environmental Monitoring and Assessment, 130, 437–454.

Zhu, B., Fitzgerald, D. G., Mayer, C. M., Rudstam, L. G., & Mills, E. L. (2006). Alteration of ecosystem function by zebra mussels in Oneida Lake: impacts on submerged macrophytes. Ecosystems, 9(6), 1017-1028.

Tables

Table 12: Summary statistics of predictor variables associated with the 482 surveys on 205 lakes included in the mixed-effect multi-level regression models.

	Variable	Count	Min	Max	Mean	Median	SD
Survey Level	Year	482	1988	2011	1998.95	1999.00	6.08
	Mean Summer Temp (°C)	482	15.82	24.98	20.59	20.33	1.59
	Degree Days	482	1478.47	3907.44	2650.83	2614.24	492.90
Lake Level	Surface Area (km2)	205	0.01	53.27	2.58	0.57	6.21
	SDI	205	1.03	8.53	2.34	2.08	1.13
	Maximum Depth (m)	205	0.91	84.12	13.49	13.11	10.41
	Latitude	205	40.61	44.73	42.20	42.10	1.19
	Longitude	205	-79.73	-71.91	-74.69	-74.13	1.58
	Elevation (m)	205	0.00	199.92	65.70	60.91	51.77
	% Deciduous Forest	205	0.00	0.93	0.26	0.20	0.27
	% Cultivated Crops	205	0.00	0.41	0.03	0.00	0.06

Table 13: Hypothesized relationships of bass population metrics with predictors from level 1 (surveys), level 2 (lakes), and level 3 (regions).

Metric	Level 1 (Survey)			Level 2 (Lake) & Level 3 (Region)									
	Year	Sum Temp	DD	Max Depth	Surf. Area	SDI	Elev	Lat	Long	% Decid	% Cultiv	Sum Temp	DD
LMB cpue	+	+	+	+/-	+/-	+	-	-	ø	+/-	+	+	+
LMB growth and condition	+	+	+	+/-	+/-	+	-	-	ø	+/-	+	+	+
SMB cpue	+	+	+	+/-	+/-	+	+/-	+/-	ø	+	+/-	+	+
SMB growth and condition	+	+	+	+/-	+/-	+	+/-	+/-	ø	+	+/-	+	+

Table 14: The number and proportion of pairwise Tukey comparisons that were significant for each bass population metric when regionalization scheme was included as the single fixed-effect predictor variable in 2-level models built from the bottom-up.

Metric	Regionalization Scheme	Total number of pairwise comparisons	Number of pairwise comparisons that were significantly different ($\alpha = 0.05$)	%
LMB CPUE	EDU	55	1	2
LMB Wr	HUC04	28	1	4
LMB LA2	ECO3	21	0	0
LMB LA5	ECO3	21	1	5
SMB CPUE	HUC04	36	5	14
SMB Wr	ECO3	15	2	13
SMB LA2	ECO3	15	4	27
SMB LA5	HUC04	28	0	0

Table 15: Results for two-level models (level 1 = survey; level 2 = lake) where region was evaluated as a fixed effect variable and model-building was done from the bottom-up. The sign of the coefficient is shown for all predictors that were retained in the final model for each bass population metric. Significant predictors are indicated by an asterisk (*). Significant interactions are identified in Table 16 and Table 17. Coefficient signs that differed from a-priori expectations are surrounded by parentheses. The "Original Model Δ AIC" column refers to the difference in AIC score between the final model (with Δ AIC of 0) and that of the original model prior to the inclusion of predictors. Note that no predictors were found to be useful in explaining variance in the model for smallmouth bass relative weight and so the original model without predictor variables had a Δ AIC of 0.

			Level 1 (Survey)			Level 2 (Lake)								
Metric	**Num. Surveys**	**Num. Lakes**	**Year**	**Sum Temp**	**DD**	**Max Depth**	**Surf. Area**	**SDI**	**Elev**	**Lat**	**Long**	**% Decid**	**% Cultiv**	**Orig. Model Δ AIC**
LMB CPUE	355	168				-*		(-)*	-*			-		30.3
LMB Wr	119	63	(-)*				+*		-				+*	23.7
LMB LA2	79	39					+*			-			(-)	8.7
LMB LA5	68	31		+*					-*	-	+*		+*	21.2
SMB CPUE	342	148	+			+*	-*	+*				+		14.7
SMB Wr	114	30												0.0
SMB LA2	66	32		+*		-*			-*				+*	14.8
SMB LA5	53	19		+*										3.2

Table 16: Coefficients for two-level models (level 1 = survey; level 2 = lake) where the response variable was a largemouth bass metric, region was evaluated as a fixed effect variable, and model-building was done from the bottom-up. The "Reduction in Total Variance (%)" column refers to the reduction in total variance between the original model which included no predictors and the final model. Coefficients for regions are not reported here.

Parameter	Level	Coef	SE	t-value	Reduction in Total Variance (%)
Relative Abundance (CPUE)					
Intercept		1.37	0.45	3.72	15
SDI	L2	-0.28	0.08	-3.69	
Max lake depth	L2	-0.01	0.00	-2.53	
% deciduous	L2	-0.43	0.37	-1.16	
Lake elevation	L2	0.00	0.00	-2.84	
Year * lake elevation	L1*L2	0.00	0.00	2.00	
Year * max lake depth	L1*L2	0.00	0.00	-1.85	
Relative Weight					
Intercept		99.46	1.48	67.16	25
Survey year	L1	-0.35	0.09	-3.70	
% cultivated	L2	38.00	11.77	3.23	
Lake elevation	L2	-0.01	0.01	-1.61	
Lake surface area	L2	0.17	0.08	2.23	
Length at Age 2					
Intercept		188.69	3.21	58.73	9
Lake surface area	L2	1.38	0.47	2.94	
Latitude	L2	-8.59	5.52	-1.56	
Length at Age 5					
Intercept		322.73	6.65	48.52	21.2
Mean summer temp	L1	30.91	9.26	3.34	
Lake elevation	L2	-0.18	0.05	-4.01	
Latitude	L2	-12.28	6.63	-1.85	
Longitude	L2	14.95	5.29	2.82	
% cultivated	L2	142.25	56.31	2.53	

Table 17: Coefficients for two-level models (level 1 = survey; level 2 = lake) where the response variable was a smallmouth bass metric, region was evaluated as a fixed effect variable, and model-building was done from the bottom-up. The "Reduction in Total Variance (%)" column refers to the reduction in total variance between the original model which included no predictors and the final model. Since no predictors improved the original model when the response variable was relative weight, that column is not relevant for the relative weight metric, hence the NA. Coefficients for regions are not reported here.

Parameter	Level	Coef	SE	t-value	Reduction in Total Variance (%)
Relative Abundance (CPUE)					
Intercept		<0.01	0.79	0.00	11
Survey year	L1	0.03	0.02	1.88	
% deciduous forest	L2	0.48	0.31	1.56	
Lake surface area	L2	-0.02	0.01	-2.37	
Max lake depth	L2	0.01	0.00	3.08	
SDI	L2	0.15	0.07	2.17	
Year * max depth	L1*L2	0.00	0.00	-1.58	
Relative Weight					
Intercept		100.97	3.26	30.95	0
Length at Age 2					
Intercept		223.04	13.48	16.55	33
Mean summer temp	L1	10.03	4.96	2.02	
Lake elevation	L2	-0.12	0.03	-3.58	
Max lake depth	L2	-0.28	0.12	-2.28	
% cultivated	L2	404.60	160.75	2.52	
Length at Age 5					
Intercept		336.94	6.72	50.14	0.05
Mean summer temp	L1	29.96	12.67	2.36	
Intercept		336.94	6.72	50.14	0.05

Table 18: Coefficients for three-level models (level 1 = survey; level 2 = lake; level 3 = region) where the response variable was a largemouth bass metric, region was evaluated as a random effect variable, and model-building was done from the bottom-up. The "Reduction in Total Variance (%)" column refers to the reduction in total variance between the original model which included no predictors and the final model.

Parameter	Level	Coef	SE	t-value	Reduction in Total Variance (%)
Relative Abundance (CPUE)					
Intercept		2.21	0.13	16.95	15
Degree days	L1	-0.00	0.00	-2.30	
Survey Year	L1	0.02	0.01	2.69	
Lake elevation	L2	-0.00	0.00	-2.89	
Max lake depth	L2	-0.00	0.00	-1.83	
SDI	L2	-0.31	0.06	-4.83	
SDI	L3	-0.57	0.26	-2.24	
Length at Age 5					
Intercept		330.11	3.45	95.82	47
Survey Year	L1	1.48	0.58	2.57	
Lake elevation	L2	-0.15	0.04	-3.98	
Lake surface area	L2	0.66	0.37	1.78	
Latitude	L3	-8.67	3.87	-2.24	
Longitude	L3	5.92	2.17	2.73	

Table 19: Results for three-level models (level 1 = survey; level 2 = lake; level 3 = region) where region was evaluated as a random effect variable and model-building was done from the bottom-up. The sign of the coefficient is shown for all predictors that were retained in the final model for each bass population metric. Significant predictors are indicated by an asterisk (*). Coefficient signs that differed from a-priori expectations are surrounded by parentheses.

				Level 1 (Survey)			Level 2 (Lake)									
Metric	Num. Surveys	Num. Lakes	Num. Regions	Year	Sum Temp	DD	Max Depth	Surf. Area	SDI	Elev	Lat	Long	% Decid	% Cultiv	Sum Temp	DD
LMB CPUE	482	205	12	+*		(-)*	-		(-)*	-*						
LMB LA5	92	40	8	+*				+		-*						

	Level 3 (Region)									
Metric	Max Depth	Surf. Area	SDI	Elev	Lat	Long	% Decid	% Cultiv	Sum Temp	DD
LMB CPUE			(-)*							
LMB LA5					-*	+*				

CHAPTER 3

Abstract

Multi-level mixed effects modeling is increasingly being used in ecological studies to explain variance in a response variable at multiple levels. As with all other modeling techniques and approaches, multi-level mixed-effects modeling requires that certain assumptions be met in order for the results to be considered reliable. In multi-level mixed-effect models there are both fixed effects and random effects and the designation of an effect as either fixed or random carries with it associated assumptions. In some cases it is not perfectly clear which designation – fixed or random – should be used to describe an effect. In this chapter, we consider the drawbacks involved with considering regionalization scheme as a fixed effect, state the reasons why regionalization scheme should not be considered to be a random effect when typical mixed-effect multi-level modeling is used, and suggest a method for incorporating a finite population correction into the likelihood equation of the model, effectively allowing regionalization scheme to be considered as a random effect which then allows the model results to be applied beyond just the regions represented by samples in the model.

Introduction

Multi-level mixed-effect modeling, whether conducted within a frequentist or Bayesian framework, is increasingly being used as a statistical technique to evaluate responses of populations and communities in spatially nested contexts for ecological studies with either observational or experimental design (Clark and Gelfand 2006, McMahon and Diez 2007). For example, Wagner et al. (2007) explored how the growth (length at age two, three) of seven different common warm and coolwater fish species varied across lakes nested within ecoregions

or watersheds in Michigan. Buckley et al. (2003) evaluated variation in vegetative and flowering stem growth, and fruit production of St John's wort (*Hypericum perforatum*) across quadrats, blocks, treatments, and sites. Helser and Lai (2004) investigated largemouth bass (*Micropterus salmoides*) growth variability across lake populations across the North American Continent. And Franklin et al. (2000) used multi-level mixed-effect modeling to partition variance in the reproductive output of northern spotted owls (*Strix occidentalis caurina*) across territories.

By definition, mixed-effect models rely on the inclusion of fixed effects as well as random effects. Fixed effects cover all possible levels of a factor whereas random effects represent only a sampling of possible levels from a larger, ideally infinite population (Patil, Sinha and Taillie 1995, Fox et al. 2004). Due to their selection at random from an assumed infinite population, random effects are more generalizable than fixed effects, but care must be taken in accurately defining a variable as either fixed or random in a multi-level mixed-effect model. For example, if all possible levels of a factor are represented by data, that factor cannot be said to have been drawn from a larger, infinite population and therefore it should not be considered a random effect.

In examining the performance of regionalization schemes for classifying lakes in relation to largemouth bass population metrics including relative abundance and individual growth in New York State in chapter one, we had to decide whether to include regionalization scheme as a random or fixed effect and there were pros and cons associated with either choice. Conceptually, the preferred approach was to include regionalization scheme as a random effect due to the generalizability of the results to regions beyond those containing actual sampling data. Including regionalization scheme as a random effect would allow for the proportion of model variance attributable to the region level of the model to be determined. A model which included a

regionalization scheme as a random variable could then be compared to a model which did not to determine the utility of the regionalization scheme in explaining variance. Likewise, models could be built using different types of regionalization schemes and the utility of each scheme in capturing among-region variance could be directly compared.

Despite the conceptual advantages of including regionalization scheme as a random effect in our multi-level mixed-effect models, in our previous work we argued that doing so was ultimately unjustifiable given the fact that the regions within the regionalization schemes were not selected from an infinite population. Practically, populations are vast enough to be considered infinite if the sample size is less than five percent of the population size (Herkenhoff and Fogli 2013), but even this looser criteria was never achieved in our work given the fact that each of the regionalization schemes we evaluated was restricted to the geographic limits of the State of New York. For example, when the response metric was largemouth bass relative abundance, the lowest proportion of sampled regions (sample size) relative to total regions (population size) was 32 percent (108 regions sampled of a total population of 339 regions; HUC10 regionalization scheme). Across all seven of the regionalization schemes we evaluated, the average proportion of sampled regions was 74 percent (± 30 percent SD) and in three cases, the proportion of sampled regions was 100 percent, meaning every region in the state was represented by data.

Ultimately, we opted to balance our desire to include regionalization scheme as a random effect in our models given the conceptual benefits with our reluctance to do so given the violation of model assumptions by including region as a fixed effect, but then accounting for unused regions by modifying the model's AIC score (Akaike 1974). For each model where regionalization scheme was included as a fixed effect, we imposed an additional penalty to the

AIC score which consisted of adding two times the number of regions not represented by data as a type of pseudo parameter equivalent in formulation to the standard penalty for estimated parameters (Equation 1). By applying this additional penalty to each model's AIC score, we in effect accounted for the full complement of regions within each scheme and justified direct comparisons of models with differing schemes. Whereas the inclusion of region as a random effect implies regions drawn from a limitless infinite population, we took the opposite tack of including region as a fixed effect and accounting for each and every one of the regions in a finite population limited geographically by the extent of the State of New York. The resulting adjustment to the model AIC scores was targeted toward the right side of the AIC equation which adds a penalty to the negative log likelihood.

$$AIC_{penalized} = -2\ln(L) + 2k + 2r \qquad \text{Equation 1}$$

where r is the number of regions with no lake data.

The problem that arose when region was considered random was that a model assumption was violated in that the sampled regions were drawn from a population of regions which was necessarily finite given their restriction to the geographic limits of the State of New York. This same problem was evident in Wagner et al.'s (2007) use of multi-level mixed effect models to evaluate variation in growth across ecoregions and watersheds in the state of Michigan. And Franklin et al (2000) encountered a similar problem when modeling northern spotted owl reproductive output in their consideration of territory as a random effect stating that "Ideally, territories should have been randomly sampled from a larger population in order to be considered random effects, but they were not." And, "…we considered territories to be a random effect, recognizing that they were not randomly drawn from a larger population."

To exemplify the problem with including region as a random effect, consider the scenario in which all regions in the scheme are populated by lakes for which response data exists. In this scenario, all regions are known and included, however the mixed-effect model assumes that there are infinitely many regions that are selected randomly and thus, erroneously, incorporates a level of variation attributable to a component that is not random. While the issue is most clearly illustrated when all regions are represented by data, an issue exists to an increasingly lesser extent even when fewer and fewer regions are represented by data.

Herein, we explore an alternative solution to this problem by incorporating a finite population correction, or fpc into the mixed-effect model likelihood equation. The *ad-hoc* solution which we presented in our previous work centered on the right side of the AIC equation. Here, we make the adjustment to the AIC equation by modifying the model likelihood on the left side of the AIC equation. While the parameter penalty we presented earlier increased the number of effective parameters in the calculation, the application of the fpc we present herein lowers the variance component associated with randomness in the regionalization schemes and approaches zero as the number of regions in a scheme approaches the total number of regions present. As such, for models where the *fpc* correction is incorporated, regionalization scheme is included as a random effect. Our objectives here were to: 1) to clearly illustrate the degree of the bias associated with treating region as a random effect without the application of a finite population correction, 2) to suggest a new method of correcting the bias associated with treating region as a random effect and 3) to compare this new method with our previously described *ad-hoc* method.

Methods

We first ran models with the fpc applied for each regionalization scheme, ranking them based on AIC score. The models we constructed and compared were each unconditional means models (which means they included no predictor variables) differing only in the particular regionalization scheme used as a random effect. We then compared the order of our fpc-corrected models to the order of models resulting from our earlier models in which region was included as a fixed effect and a subsequent penalized-AIC criterion was used to compare models.

We defined the *fpc* as

$$fpc = \frac{N - n}{N - 1} \qquad \text{Equation 2}$$

where *N* is the total number of regions in a regionalization scheme and *n* is the number of regions represented by data (Isserlis 1918, Shaw and Goldstein 2012).

We incorporated the *fpc* into the standard mixed-effect likelihood equation provided by Verbeke and Moelnberhs (2000) and Zuur et al. (2009) (Equation 3). Specifically, the *fpc* was inserted just before the mixed-effects variance term on the left-hand side of equation 3. Since we were comparing models which differed only in their random component, we used the restricted maximum likelihood (REML) mixed-effect likelihood equation rather than the maximum likelihood mixed-effect likelihood equation because the maximum likelihood equation should only be used to compare models which differ in their fixed effects, but not their random effects (Zuur et al. 2009).

$$L_{REML}(\theta) = C * \left| fpc * \sum_{g=1}^{G} X_g' \times V_g^{-1} \times X_g \right|^{-0.5} * L_{ML}(\theta)$$ Equation 3

$$L_{ML}(\theta) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} * exp\left(- \frac{1}{2\sigma^2} \sum_{i}^{n} (x_i - \mu)^2 \right)$$

where…

Symbol	Definition
C	Constant
g	Index of group
G	Total number of groups in the dataset
X	Vector of data values
V	Variance
π	pi
σ^2	Variance (a.k.a., "sigma squared")
n	Number of residuals in simple linear regression
i	Index of residual
x	Value of residual
μ	Mean of all residuals

For each of the seven regionalization schemes considered, we ran two mixed-effect

models, one which did not include the *fpc* and another which did. We then computed a negative log likelihood value and an AIC score for each model, noting differences in these scores between *fpc*-corrected models and non-corrected models. We then ranked regionalization schemes in order from best performing to worst performing based on their *fpc*-corrected AIC scores (where a lower AIC score indicates a better performing model).

Finally, we compared the ranking of regionalization schemes obtained using the *fpc* method to our earlier scheme rankings derived from further penalizing the AIC scores. A distinct series of model comparisons was conducted for each of the two largemouth bass population metrics considered, namely relative abundance and length at age five.

Results

The AIC scores for the models with and without the *fpc* correction are compared in Table 20 where the response variable was largemouth bass relative abundance and in Table 21 where the response variable was largemouth bass length at age five. Spearman rank correlations revealed that the AIC rankings were positively (but not significantly) correlated when the response variable was largemouth relative abundance (rho = 0.54, p-value = 0.22), and significantly positively correlated when the response variable was largemouth bass length at age 5 (rho = 0.96, p-value = <0.001.

Using the penalized AIC method, we had determined previously that the performance of regionalization schemes in clustering lakes for largemouth bass relative abundance should be ranked (from best to worst-performing) as follows: EDU, ECO4, HUC04, HUC08, ECO3, HEX, and HUC10 (chapter one). Using the *fpc*-corrected method, we found that while the ranking of regionalization schemes was not exactly the same, the order was comparable (Table 22). Three

(EDU, HUC04, and HUC10) of the seven schemes held the same rank in both cases. The schemes which ranked in the top four (EDU, ECO4, HUC04, and HUC08) were the same under both methods. And, there was symmetry in some cases where schemes ranked differently across the two methods. For example, the ECO4 scheme ranked second under the penalized AIC method and fourth under the *fpc*-corrected method and the HUC08 scheme ranked fourth under the *fpc*-corrected method and second under the penalized AIC method. Also, the ECO3 scheme ranked fifth under the penalized AIC method and sixth under the *fpc*-corrected method and the HEX scheme ranked sixth under the penalized AIC method and fifth under the *fpc*-corrected method (Table 22). Spearman rank correlations revealed that the AIC rankings were significantly positively correlated when the response variable was largemouth bass relative abundance (rho = 0.82, p-value = 0.02).

Similarly, when the response variable was largemouth bass length at age 5, the ranking of schemes under the penalized-AIC method was (from best to worst-performing) ECO3, EDU, HUC04, HUC08, ECO4, HEX, and HUC10 (chapter one). And after incorporating the *fpc*, the ordering of regionalization schemes was similar (Table 23). Again, the ranking was exactly the same across the two methods for three schemes (EDU, HUC08, and HUC10). However, one notable dissimilarity in the ranking of schemes between the two methods was the position of the HEX scheme which ranked sixth under the penalized AIC method, but ranked first under the *fpc*-corrected method (Table 23). Spearman rank correlations revealed that the AIC rankings were positively (but not significantly) correlated when the response variable was largemouth bass length at age 5 (rho = 0.16, p-value = 0.73).

Discussion

Meeting the assumptions of statistical models is a fundamental necessity for acquiring reliable results from their use. The increasing use of multi-level mixed-effect models in ecology should be paralleled by continual consideration of whether model assumptions are being met. In cases where assumptions are not being met, it is useful to consider alternative model designs which make use of available data, constrained as they may be, while also satisfying model assumptions. We have described two methods for dealing with the problem where factor levels in a multi-level mixed-effect model are not drawn from an infinite population which is a model assumption violation if the levels are included as random effects: 1) a penalized AIC method (chapter one) and 2) the fpc-correction method presented herein.

In the case of largemouth bass relative abundance, the fpc-correction method clearly produces results which differ from the model outputs generated with no correction applied (Table 20) illustrating the bias embedded in models which do not meet model assumptions. However, this was less apparent in the case of the largemouth bass length age 5 (Table 21).

The results we obtained when we ranked regionalization schemes based on the *fpc* correction were similar to the results we obtained when we ranked regionalization schemes by further penalizing the AIC scores based on the number of regions without lake data, especially in the case of largemouth bass relative abundance. This similarity in rankings regardless of method used suggests that the AIC-penalizing method, which is simpler and perhaps easier to implement is as viable a method for dealing with the problem (of a variable which is included as a random effect yet has a finite number of levels) when compared with the quantitatively more involved *fpc*-correction method.

The two methods we implemented to address the problem of having a random variable

drawn from a finite set of options dealt with different components of the basic AIC formula. The method which added a further penalty to the AIC score beyond the penalty already imposed for the number of parameters in the model dealt with the right-hand side of the AIC formula and focused on the number of un-occupied regions. That the application of each of the two approaches produced similar results is reassuring.

While we applied these two methods for dealing with a random variable drawn from a finite set of options to a situation involving largemouth bass population metrics (relative abundance and growth) summarized at a lake and region level, these model corrections could foreseeably have broad applicability in a wide range of ecological studies. In any case where the assumption of a random variable is violated by severe restrictions in the number of levels available in the variable, either of these two corrective measures could be implemented and our results suggest that the simpler method of penalizing model AIC scores with the number of unused levels performs similarly to the more quantitatively satisfying method of inserting a correction for finite populations into the mixed effect likelihood equation.

Researchers may be reluctant to take advantage of powerful modern statistical approaches if they recognize that their data do not meet certain model assumptions and do not know how to adjust their approaches so that assumptions are not violated. Worse, researchers may be tempted to use a model even if their data do not meet model assumptions if no clear alternative is apparent. The methods and results described herein provide alternative ways of dealing with the problem of including factor levels as random effects when they are actually not selected from infinite populations. It is our hope that the solutions described herein will promote the appropriate use of new statistical approaches while simultaneously discouraging the inappropriate application of such approaches.

References

Buckley, Y. M., Briese, D. T., & Rees, M. (2003). Demography and management of the invasive plant species Hypericum perforatum. I. Using multi-level mixed-effects models for characterizing growth, survival and fecundity in a long-term data set. Journal of Applied Ecology, 40(3), 481-493.

Clark, J.S. & Gelfand, A.E. (2006). Hierarchical Modeling for the Environmental Sciences: Statistical Methods and Applications. Oxford University Press, Oxford.

Fox, G.A., S Negrete-Yankelevich, and V.J. Sosa. 2015, Ecological Statistics: Contemporary theory and application. OUP Oxford. pp 311-315.

Franklin, A. B., Anderson, D. R., Gutiérrez, R. J., & Burnham, K. P. (2000). Climate, habitat quality, and fitness in northern spotted owl populations in northwestern California. Ecological Monographs, 70(4), 539-590.

Herkenhoff, L., & Fogli, J. (2013). Applied Statistics for Business and Management Using Microsoft Excel. Springer. pp 148.

Isserlis, L. (1918). On the Value of a Mean as Calculated from a Sample. Journal of the Royal Statistical Society, 75-81.

McMahon, S. M., & Diez, J. M. (2007). Scales of association: hierarchical linear models and the measurement of ecological systems. Ecology Letters, 10(6), 437-452.

Patil, G.P., A.K. Sinha, and C. Taillie. 1995. Finite Population Corrections for ranked set sampling. Annuls of the Institute of Statistical Mathematics. 47. 621-636.

Shaw, S. C. and Goldstein, M. (2012) Finite population corrections for multivariate Bayes sampling. Journal of Statistical Planning and Inference, 142 (10). pp. 2844-2861.

Singer, Judith D. 1998. Using SAS PROC MIXED to fit multilevel models, hierarchical models and individual growth models. Journal of Educational and Behavioral Statistics. 24. 323-355

Stan Development Team. 2015. RStan: the R interface to Stan, Version 2.7.0 http://mc-stan.org/interfaces/rstan.html.

Verbeke, G and G. Molenberghs. 2000. Linear Mixed Models for Longitudinal Data. Springer Series in Statistics. Springer. page 45 (equation 5.7)

Zuur, A.F. E.N. Ieno, N.J. Walker, A.A. Saveliev, and G.M. Smith. 2009. Mixed Effects Models and Extensions in Ecology with R. Springer. page 121.

Tables

Table 20 Comparison of uncorrected and fpc-corrected negative log likelihood and AIC scores when the response metric was largemouth bass relative abundance

	Uncorrected		fpc-corrected		AIC difference	uncorrected rank	fpc-corrected rank
Scheme	-lnL	AIC	-lnL	AIC			
EDU	308.53	623.06	300.74	607.48	6.68	4	1
ECO3	316.87	639.73	308.86	623.71	6.87	7	6
ECO4	307.89	621.79	307.16	620.32	1.46	3	4
HEX	307.83	621.66	307.61	621.23	0.43	2	5
HUC04	312.59	631.17	304.70	615.39	6.76	5	3
HUC08	304.85	615.70	304.21	614.41	1.29	1	2
HUC10	313.61	633.22	313.43	632.85	0.37	6	7

Table 21 Comparison of uncorrected and fpc-corrected negative log likelihood and AIC scores when the response metric was largemouth bass length at age 5

	uncorrected		fpc-corrected		AIC difference	uncorrected rank	fpc-corrected rank
Scheme	-lnL	AIC	-lnL	AIC			
EDU	225.90	457.80	225.42	456.84	-0.96	2	2
ECO3	226.74	459.47	226.08	458.15	-1.32	3	3
ECO4	227.81	461.62	227.61	461.23	-0.39	5	6
HEX	225.33	456.66	225.28	456.55	-0.11	1	1
HUC04	227.82	461.63	227.10	460.19	-1.44	6	5
HUC08	226.95	459.90	226.73	459.46	-0.43	4	4
HUC10	228.11	462.22	228.07	462.14	-0.08	7	7

Table 22 Rankings of schemes when the response metric is largemouth bass relative abundance

Scheme	fpc-corrected	penalized AIC
EDU	1	1
ECO3	6	5
ECO4	4	2
HEX	5	6
HUC04	3	3
HUC08	2	4
HUC10	7	7

Table 23 Rankings of schemes when the response metric is largemouth bass length at age 5

Scheme	fpc-corrected	penalized AIC
EDU	2	2
ECO3	3	1
ECO4	6	5
HEX	1	6
HUC04	5	3

Scheme	fpc-corrected	penalized AIC
HUC08	4	4
HUC10	7	7

APPENDIX A: CHAPTER TWO SUPPLEMENTARY MATERIALS

Top-down approach

Zuur et al. (2009) and Grueber et al. (2011) advocate a top-down approach to mixed-effects model-building. In our paper, we elected to use a bottom-up approach because of its clear emphasis on hypothesized biological relationships between the environmental covariates and the response metrics. However, to provide a comparison of the two approaches, we also ran the models using the top-down approach and have presented our methods and results here as supplementary material.

Methods

Two-Level Top-Down Models:

Following the top-down strategy recommended by Zuur et al. (2009), we began by setting up a full regression model – one in which all predictor variables were included – for each bass response metric. For each response metric, we first explored three options for the optimal random structure of the model: 1) no random structure, 2) random intercepts only, and 3) random intercepts and slopes. We did this by running three separate models and then comparing AIC scores. A model was selected as best if its delta AIC value was at least 2 units lower than the other models. In cases where the difference in delta AIC values between the model with no random structure and the model with random intercepts was less than 2 units indicating the equivalency of either model, we opted to include random intercepts rather than have no random structure. As suggested by Zuur et al. (2009), we used the Restricted Maximum Likelihood (REML) method to compare these three models which all had the same fixed variable structure, and differed only in their random variable structure.

Once the optimal random structure of the full model was established, we turned our

attention to identifying the optimal fixed structure. Here, we included all cluster-centered predictor variables in a full model and if the bottom-up approach had revealed any important interactions, we also included those interactions in the full model. Then, we relied on AIC scores to compare the full model against subsequent reduced models consisting of every possible combination of cluster-centered predictor variables. We used the dredge function implemented in the MuMIn package in R (Bartoń, 2014) to generate all sub models. Since we were exploring alternative combinations of fixed effects in these mixed-effects models and the random effect, lake was kept constant; we used the Maximum Likelihood (ML) method rather than Restricted Maximum Likelihood (REML) method to compare models as suggested by Zuur et al. (2009).

All of the regression models that were within 2 delta AIC units of the top model were considered equally optimal in explaining the variance associated with the response variable. Therefore, following the example of Grueber et al. (2011), we performed model averaging on all the models within 2 delta AIC units of the top model to acquire final predictor coefficients and 95% confidence intervals associated with them. Following model averaging, an importance value was assigned to every environmental variable that was represented at least once in the models with delta AIC units ≤ 2. The importance value was defined as the sum of the AIC weights (recalculated for all models with delta AIC units ≤ 2) which included the particular environmental variable. The importance values ranged from 0.0 to 1.0. For example, if a particular environmental variable was included in all the top models, the importance value of that variable would be equal to 1.0.

In summary, a distinct series of regression analyses, AIC comparisons and model averages were run for each of the eight bass population metrics. All models were run in Program R (R Core Team 2014) using the lme4 (Bates et al. 2014) or nlme package (Pinheiro et al. 2014).

Both packages were required because only the nlme package could be used to run models with no random structure, but we preferred the model syntax of the lme4 package for the construction of the large full models.

Results

All 30 of the response-predictor relationships seen in the 2-level bottom-up models (Table 4 in main article) were also observed in the 2-level top-down models (*Table 24*) and in every case, the sign of the relationship matched between the bottom-up and top-down models (Table 4, *Table 24*). Of the 22 relationships that were significant in the bottom-up models, slightly more than two-thirds (15, 68%) were also significant in the top-down models. While the top-down models (*Table 24*) included 46 more response-predictor relationships than the bottom-up models (Table 4), none of those additional relationships were significant in the top-down models. In other words, it was never the case that there were significant predictors in the top-down models that were not significant in the bottom-up models. Coefficients of all models built using the top-down strategy are provided in *Table 25* (for largemouth bass) and *Table 26* (for smallmouth bass).

One second level predictor variable, percent cultivated crops in the lake catchment area, was present in all eight of the final 2-level top-down models (*Table 24*). Four second level predictor variables (lake surface area, lake elevation, longitude and mean summer temperature) were each present in seven out of the eight 2-level top-down models, the exception in all four cases being smallmouth bass length at age 5 (*Table 24*). And three predictors (survey year, latitude, and degree days summarized at the lake level) were present in six of the eight 2-level top-down models (*Table 24*).

Across all eight population metrics, the total number of models with delta AIC-values ≤ 2 ranged between 5 (for smallmouth bass length at age 5) and 37 (for smallmouth bass length at age 2) with a mean (±SD) of 14.50 (± 10.27). Due to software limitations, model averaging could only be done on thirty or fewer models. The smallmouth bass length at age 2 metric was the only one for which > 30 models had delta AIC values ≤ 2, so in that case, we model-averaged only the thirty models which had the lowest AIC scores (the highest delta AIC value being 1.68). For the other 7 metrics, we model-averaged all models with delta AIC values ≤ 2. The importance scores generated for each predictor following model-averaging are presented in *Table 27*. For all metrics except smallmouth bass relative weight, there was at least one predictor variable with an importance score of 1.0 meaning that the predictor was present in all the best models which were included in model averaging. Two predictors (lake surface area and survey year) each had importance scores of 1.0 in four of the eight model average results and the predictor, lake elevation had an importance score of 1.0 in three of the eight model average results (*Table 27*). Of the 22 significant predictors in the 2-level bottom-up models, 17 (77%) had importance values of 1.0 in the 2-level top-down models.

References

Bartoń, K. (2014). MuMIn: Multi-model inference. R package version 1.9.13. http://CRAN.R-project.org/package=MuMIn.

Bates, D., M. Maechler, B. Bolker and S. Walker (2014). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.0-5. http://CRAN.R-project.org/package=lme4.

Burnham, K. P. and Anderson, D. R. (2002) Model selection and multimodel inference: a practical information-theoretic approach. 2nd ed. New York, Springer-Verlag.

Grueber, C.E., S. Nakagawa, R.J. Laws, and I.G. Jamieson (2011). Multimodel inference in ecology and evolution: challenges and solutions. Journal of Evolutionary Biology, 24, 699-711.

Idrisi, Nasseer, Edward L. Mills, Lars G. Rudstam, and Donald J. Stewart. "Impact of zebra mussels (Dreissena polymorpha) on the pelagic lower trophic levels of Oneida Lake, New York." Canadian Journal of Fisheries and Aquatic Sciences 58, no. 7 (2001): 1430-1441.

Pinheiro J., D. Bates, S. DebRoy, D. Sarkar, and R Core Team (2014). nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-117, URL: http://CRAN.R-project.org/package=nlme.

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Walther, Gian-Reto, Eric Post, Peter Convey, Annette Menzel, Camille Parmesan, Trevor JC Beebee, Jean-Marc Fromentin, Ove Hoegh-Guldberg, and Franz Bairlein. "Ecological responses to recent climate change." Nature 416, no. 6879 (2002): 389-395.

Zuur, A.F., E.N. Ieno, N.J. Walker, A.A. Saveliev, and G.M. Smith (2009). Mixed effects models and extensions in ecology with R. Springer, New York, NY. pp 116-122.

Zweifel, R. D., Hayward, R. S., & Rabeni, C. F. (1999). Bioenergetics insight into black bass distribution shifts in Ozark border region streams. North American Journal of Fisheries Management, 19, 192–197.

Tables

Table 24: Results for two-level models (level 1 = survey; level 2 = lake) where region was evaluated as a fixed effect variable and model-building was done from the top-down. The sign of the coefficient is shown for all predictors that were retained in the final model for each bass population metric. Significant predictors are indicated by an asterisk (*). Coefficient signs that differed from a-priori expectations are surrounded by parentheses.

				Level 1 (Survey)			Level 2 (Lake)									
Metric	**Num. Surveys**	**Num. Lakes**	**Region Included**	**Year**	**Sum Temp**	**DD**	**Max Depth**	**Surf. Area**	**SDI**	**Elev**	**Lat**	**Long**	**% Decid**	**% Cultiv**	**Sum Temp**	**DD**
LMB CPUE	482	205	**YES**	+		+	-	+	(-)*	-*		-		(-)	+	-
LMB Wr	119	63	**NO**	(-)	(-)	+		+*		-	(+)	-	+	+*	(-)	
LMB LA2	84	43	**NO**			+		+*		-				(-)	+	+
LMB LA5	92	40	**YES**	+*	(-)	(-)		+*		-*	-	+*		+*	(-)	+
SMB CPUE	455	194	**YES**	+			+*	-	+*	+	+	+	+	+	+	
SMB Wr	114	30	**YES**	+	(-)		-	+	+	-	+	-		+	+	+
SMB LA2	76	37	**YES**		(-)		-	+	-	-*	+	-	(-)	+	(-)	(-)
SMB LA5	64	24	**NO**	+*		(-)					-	-	+	+		(-)

Table 25: Coefficients for two-level models (level 1 = survey; level 2 = lake) where the response variable was a largemouth bass metric, region was evaluated as a fixed effect variable, and model-building was done from the top-down. Coefficients for regions are not reported here.

Parameter	Level	Coef	Adjusted SE	z-value	Pr(>|z|)
Relative Abundance (CPUE)					
Intercept		-9.75	23.94	0.41	0.68
Degree days	L1	0.00	0.00	1.11	0.27
Survey year	L1	0.01	0.01	0.49	0.62
% cultivated	L2	-0.06	0.44	0.13	0.90
Degree days	L2	0.00	0.00	1.31	0.19
Lake elevation	L2	0.00	0.00	2.45	0.01
Lake surface area	L2	0.01	0.01	0.69	0.49
Longitude	L2	-0.01	0.06	0.22	0.83
Max lake depth	L2	0.00	0.00	1.57	0.12
Mean summer temp	L2	0.00	0.01	0.19	0.85
SDI	L2	-0.33	0.07	4.85	0.00
Relative Weight					
Intercept		99.69	0.63	159.10	0.00
Degree days	L1	0.00	0.00	0.17	0.86
Mean summer temp	L1	-0.15	0.16	0.98	0.33
Survey year	L1	-0.40	0.10	3.78	0.00
% cultivated	L2	35.90	10.17	3.53	0.00
Degree days	L2	0.00	0.00	0.53	0.60
Lake elevation	L2	-0.01	0.01	1.48	0.14
Lake surface area	L2	0.19	0.08	2.47	0.01
Latitude	L2	0.22	0.61	0.37	0.71
Longitude	L2	-0.05	0.22	0.24	0.81
Mean summer temp	L2	-0.01	0.04	0.22	0.83
Length at Age 2					
Intercept		189.23	3.44	55.01	0.00
Degree days	L1	0.00	0.01	0.18	0.85
% cultivated	L2	-17.11	36.82	0.46	0.64
Degree days	L2	0.00	0.01	0.41	0.68
Lake elevation	L2	0.00	0.01	0.28	0.78
Lake surface area	L2	0.77	0.35	2.23	0.03
Mean summer temp	L2	0.15	0.63	0.24	0.81
Length at Age 5					
Intercept		328.52	9.92	33.10	0.00
Degree days	L1	0.00	0.02	0.11	0.91
Mean summer temp	L1	-1.14	2.35	0.49	0.63
Survey year	L1	1.59	0.58	2.73	0.01
% cultivated	L2	140.01	57.75	2.42	0.02

Parameter	Level	Coef	Adjusted SE	z-value	Pr(>|z|)
Degree days	L2	0.02	0.02	0.98	0.33
Lake elevation	L2	-0.14	0.04	3.13	0.00
Lake surface area	L2	0.77	0.33	2.32	0.02
Latitude	L2	-10.46	8.87	1.18	0.24
Longitude	L2	10.96	5.57	1.97	0.05
Mean summer temp	L2	-2.68	2.80	0.95	0.34

Table 26: Coefficients for two-level models (level 1 = survey; level 2 = lake) where the response variable was a smallmouth bass metric, region was evaluated as a fixed effect variable, and model-building was done from the top-down. Coefficients for regions are not reported here.

Parameter	Level	Coef	Adjusted SE	z-value	Pr(>|z|)
Relative Abundance (CPUE)					
Intercept		0.27	0.80	0.34	0.74
Survey year	L1	0.01	0.01	1.26	0.21
% cultivated	L2	0.15	0.62	0.23	0.82
% deciduous forest	L2	0.50	0.32	1.58	0.11
Lake elevation	L2	0.00	0.00	0.39	0.70
Lake surface area	L2	-0.02	0.01	2.81	0.00
Latitude	L2	0.16	0.18	0.87	0.38
Longitude	L2	0.19	0.13	1.49	0.14
Max lake depth	L2	0.01	0.00	3.95	0.00
Mean summer temp	L2	0.00	0.01	0.14	0.89
SDI	L2	0.15	0.06	2.40	0.02
Survey year * Longitude	L1, L2	-0.01	0.00	3.02	0.00
Relative Weight					
Intercept		93.48	4.65	20.11	0.00
Mean summer temp	L1	-0.08	0.26	0.32	0.75
Survey year	L1	0.13	0.20	0.67	0.50
% cultivated	L2	1.40	8.42	0.17	0.87
Degree days	L2	0.00	0.00	0.12	0.91
Lake elevation	L2	0.00	0.00	0.13	0.90
Lake surface area	L2	0.27	0.19	1.45	0.15
Latitude	L2	1.67	2.15	0.78	0.44
Longitude	L2	-0.67	0.87	0.77	0.44
Max lake depth	L2	-0.03	0.04	0.61	0.54
Mean summer temp	L2	0.08	0.25	0.34	0.74
SDI	L2	0.08	0.39	0.21	0.83
Length at Age 2					
Intercept		205.00	18.57	11.04	0.00
Mean summer temp	L1	-1.16	2.34	0.49	0.62
Lake surface area	L2	0.34	0.90	0.38	0.70
Degree days	L2	0.00	0.00	0.21	0.84

Parameter	Level	Coef	Adjusted SE	z-value	Pr(>|z|)
Lake elevation	L2	-0.15	0.04	3.36	0.00
Latitude	L2	2.49	4.55	0.55	0.58
Longitude	L2	-2.24	4.17	0.54	0.59
Max lake depth	L2	-0.07	0.14	0.53	0.59
% deciduous forest	L2	-4.04	15.03	0.27	0.79
% cultivated	L2	76.37	145.26	0.53	0.60
SDI	L2	-1.49	3.99	0.37	0.71
Mean summer temp	L2	-0.07	0.49	0.15	0.88
Length at Age 5					
Intercept		329.31	4.84	68.00	0.00
Degree days	L1	-0.04	0.02	1.63	0.10
Latitude	L2	-14.27	8.46	1.69	0.09
% cultivated	L2	204.74	145.50	1.41	0.16
Survey year	L1	3.46	0.94	3.67	0.00
Longitude	L2	-1.46	2.99	0.49	0.63
% deciduous forest	L2	14.98	29.97	0.50	0.62
Degree days	L2	0.00	0.01	0.34	0.73

Table 27: Model-average importance scores of predictors in two-level models (level 1 = survey; level 2 = lake) built from the top-down. The "Count" column refers to how many of the models that were averaged included the predictor.

Largemouth bass

Relative Abundance (CPUE)

Predictor	Level	Import.	Count
EDU		1.00	19
Lake elevation	L2	1.00	19
SDI	L2	1.00	19
Max lake depth	L2	0.90	17
Degree days	L2	0.81	15
Degree days	L1	0.70	12
Lake surface area	L2	0.47	9
Survey Year	L1	0.31	6
Longitude	L2	0.10	2
Mean summer temp	L2	0.09	2
% cultivated	L2	0.07	2

Relative Weight

Lake surface area	L2	1.00	12
% cultivated	L2	1.00	12

Smallmouth bass

Relative Abundance (CPUE)

Predictor	Level	Import.	Count
Lake surface area	L2	1.00	8
Longitude	L2	1.00	8
Max lake depth	L2	1.00	8
SDI	L2	1.00	8
Survey Year	L1	1.00	8
Survey Year * Longitude	L1*L2	1.00	8
% deciduous forest	L2	0.89	7
HUC04		0.88	7
Latitude	L2	0.61	5
Lake elevation	L2	0.21	2
% cultivated	L2	0.17	2
Mean summer temp	L2	0.08	1

Relative Weight

Lake surface area	L2	0.82	23

Largemouth bass

Survey Year	L1	1.00	12
Lake elevation	L2	0.87	10
Mean summer temp	L1	0.66	8
Degree days	L2	0.35	5
Latitude	L2	0.21	3
Longitude	L2	0.13	2
Mean summer temp	L2	0.08	1
Degree days	L1	0.07	1

Length at Age 2

Lake surface area	L2	1.00	7
% cultivated	L2	0.28	2
Degree days	L2	0.25	2
Lake elevation	L2	0.13	1
Mean summer temp	L2	0.11	1
Degree days	L1	0.09	1

Length at Age 5

Lake surface area	L2	1.00	6
Lake elevation	L2	1.00	6
Longitude	L2	1.00	6
% cultivated	L2	1.00	6
Survey Year	L1	1.00	6
Latitude	L2	0.67	4
Mean summer temp	L2	0.62	4
Degree days	L2	0.61	3
ECO3		0.33	2
Mean summer temp	L1	0.22	1
Degree days	L1	0.21	2

Smallmouth bass

Longitude	L2	0.51	15
Latitude	L2	0.51	14
Survey Year	L1	0.45	13
Max lake depth	L2	0.39	11
Mean summer temp	L2	0.13	5
ECO3		0.11	4
Mean summer temp	L1	0.11	4
SDI	L2	0.06	2
% cultivated	L2	0.05	2
Degree days	L2	0.04	2
Lake elevation	L2	0.04	2

Length at Age 2

Lake elevation	L2	1.00	30
ECO3		0.92	27
Latitude	L2	0.31	10
Longitude	L2	0.30	9
Max lake depth	L2	0.29	8
Mean summer temp	L1	0.29	9
% cultivated	L2	0.29	9
Lake surface area	L2	0.17	6
SDI	L2	0.17	6
% deciduous forest	L2	0.10	3
Degree days	L2	0.06	2
Mean summer temp	L2	0.03	1

Length at Age 5

Survey Year	L1	1.00	5
Latitude	L2	0.87	4
Degree days	L1	0.86	4
% cultivated	L2	0.77	4
% deciduous forest	L2	0.30	2
Longitude	L2	0.23	1
Degree days	L2	0.14	1